# TOWARDS A SEMANTIC WEB OF THINGS FOR SMART CITIES

PhD Thesis

2017

Shaun Howell

**SUMMARY**

Realising the value of the growing quantity of web-enabled devices and data is a significant global challenge, and is essential in overcoming the mounting global environmental and economic issues. This is especially true in urban environments, where the potential to leverage technology for operational performance improvements is highest, due to the high density of many interlinked systems. This thesis hypothesises that moving beyond the state-of-the-art of Internet of Things technologies, to a Semantic Web of Things approach, can improve the outcomes of technology interventions for stakeholders, by improving application-layer interoperability. The premise is that by providing a rich and shared understanding of the cyber-physical context of devices, services, and data, applications are able to interoperate better. This in turn leads to a more integrated consideration of the problem space by business services, and so a more holistic optimisation can be achieved, across previously siloed systems.

The methodology adopted was an iterative experimentation process alongside experts, culminating in the development of a Semantic Web of Things platform for smart cities. This consists of an integrated suite of APIs for accessing semantically-enriched built-environment data from various perspectives. This includes an IoT resource discovery endpoint which extracts semantic metadata from a triple store and transforms it to be compliant with the recent Hypercat standard. The API also exposes a full SPARQL endpoint for rich querying of the data, as well as BIM, CityGML, and timeseries endpoints for accessing specific views of the data. To further promote resource discovery and interoperability, the platform includes a 3D GUI for visually exploring the city, building, and sensor data, and is built on a comprehensive smart city ontology which extends the recent BSI smart city ontology and aligns this with several relevant de facto standards.

To support the final design science stage and provide a rigorous exploration of the hypothesis, participatory action research methods were iteratively undertaken across 6 research projects, involving engagement with circa 40 organisations. This work considered the sub-domains of smart cities, and initially focused on the energy domain. Software and ontologies were developed and analysed alongside experts, before an extended learning iteration in the water domain was undertaken, producing a smart water semantic model and platform.

The work demonstrates that a Semantic Web of Things approach does improve application-layer interoperability. Some of the results observed through this project include i) reducing energy consumption in public buildings by circa 30% through a smart retrofit BEMS, ii) enabling water utilities to better manage regulatory compliance and network management, and iii) maximising the profitability of domestic renewable energy generation through smart holonic microgrids.

Semantic technologies are well suited to addressing the 'variety' of big data in IoT systems, and support a system of systems approach to smart city management. Whilst existing research in this area focuses on annotating sensors with ICT descriptors, this work shows that integrating this with rich domain context is beneficial in promoting interoperability and discoverability. Future work involves investigating the use of the artefacts developed in other smart city domains, and furthering the consensus-building process towards standardisation.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ACRONYMS

| Acronym | Meaning |
|---|---|
| **AC** | Alternating current |
| **ACL** | Agent communication language |
| **ADE** | Application domain extension |
| **ADR** | Automated demand response |
| **AEC** | Architecture, Engineering, Construction |
| **AECFM** | Architecture, engineering, construction, & facility management |
| **AI** | Artificial intelligence |
| **AIOTI** | Alliance for Internet of Things Innovation |
| **AJAX** | Asynchronous JavaScript and XML |
| **AM** | Asset management |
| **AMQP** | Advanced Message Queuing Protocol |
| **ANN** | Artificial neural network |
| **API** | Application programming interface |
| **BAS** | Building automation system |
| **BCN** | Barcelona |
| **BEMS** | Building energy management system |
| **BI** | Business intelligence |
| **BIM** | Building information modelling |
| **BMS** | Building management system |
| **BPMN** | Business process modelling notation |
| **BRE** | Building Research Establishment |
| **BSI** | British Standards Institution |
| **CAD** | Computer Aided Design |
| **CAPEX** | Capital Expenditure |
| **CCLA** | City climate leadership awards |
| **CEN** | Comité Européen de Normalisation |
| **CHP** | Combined heat and power |
| **CIM** | Common information model |
| **CoAP** | Constrained Application protocol |
| **CPU** | Central Processing Unit |
| **CRUD** | Create, read, update, delete |
| **CSO** | Combined sewer overflow |
| **CSS** | Custom style sheet |
| **CSV** | Comma separated values |
| **CUSP** | Computational Urban Sustainability Platform |
| **DB** | Database |
| **DER** | Distributed energy resources |
| **DG** | Distributed generation |
| **DHN** | District heating network |
| **DL** | Description logic |
| **DSM** | Demand-side management |
| **DSO** | Distribution service operator |

| | |
|---|---|
| **DST** | Decision support tool |
| **EAI** | Enterprise application integration |
| **EIP** | European innovation partnership |
| **ESB** | Enterprise service bus |
| **ETL** | Extract, transform, load |
| **FIPA** | Federation of Intelligent Physical Agents |
| **FM** | Facility manager or facilities management |
| **GA** | Genetic algorithm |
| **GIS** | Geographic information systems |
| **GML** | Geography markup language |
| **GUI** | Graphical user interface |
| **GUID** | Globally unique identifier |
| **HEMS** | Home energy management system |
| **HMAS** | Holonic multi-agent system |
| **HTML** | Hypertext markup language |
| **HTTP** | Hypertext transfer protocol |
| **IBM** | International Business Machines |
| **ICT** | Information and communication technologies |
| **ID** | Identifier |
| **IFC** | Industry foundation classes |
| **IIoT** | Industrial Internet of Things |
| **IoT** | Internet of Things |
| **IP** | Internet Protocol |
| **IPID** | Individual pipe identifier |
| **IRI** | Internationalized Resource Identifier |
| **IS** | Information system |
| **JSON** | JavaScript Object Notation |
| **KB** | Knowledge base |
| **KBS** | Knowledge-based system |
| **KnoHolEM** | Knowledge-based holistic energy management of public buildings |
| **KPI** | Key performance indicator |
| **LD** | Linked data |
| **MAS** | Multi-agent system |
| **MES** | Multi-energy system |
| **MQTT** | Message queue telemetry transport |
| **MRA** | Multi-regression analysis |
| **MVD** | Model-view definition |
| **OOP** | Object-oriented programming |
| **OOPS** | Ontology pitfall scanner |
| **OPEX** | Operational expenditure |
| **OWL** | Web ontology language |
| **PAS** | Publicly available specification |
| **PCA** | Principal component analysis |
| **PEV** | Plugin electric vehicle |
| **PHP** | PHP Hypertext Pre-processor |
| **PID** | proportional–integral–derivative |

| | |
|---|---|
| **PMV** | Predicted mean vote |
| **PV** | Photovoltaic |
| **QoS** | Quality of Service |
| **RAML** | RESTful API Modelling Language |
| **RDF** | Resource description framework |
| **RDFS** | Resource description framework schema |
| **RES** | Renewable energy source |
| **REST** | Representational state transfer |
| **ROI** | Return on Investment |
| **RTC** | Real-time controller |
| **SCADA** | Supervisory control and data acquisition |
| **SCCM** | Smart city concept model |
| **SCO** | Smart city ontology |
| **SDO** | Standards development organisation |
| **SG** | Smart grid |
| **SGIM** | Smart grid information model |
| **SOA** | Service-oriented architecture |
| **SPARQL** | SPARQL Protocol and RDF Query Language |
| **SQL** | Structured Query Language |
| **SSN** | Semantic sensor network |
| **STEEP** | Social, technological, environmental, economic, political |
| **STS** | Socio-technical system |
| **SWIM** | Semantic Water Information Model |
| **SWoT** | Semantic Web of Things |
| **SWRL** | Semantic Web Rule Language |
| **TD** | Thing Description |
| **TLS** | Transport Layer Security |
| **UML** | Unified Modelling language |
| **URI** | Uniform Resource Identifier |
| **URL** | Uniform Resource Locator |
| **VPP** | Virtual power plant |
| **WCIM** | Water catchment information model |
| **WebGL** | Web Graphics Library |
| **WISDOM** | Water Analytics and Intelligent Sensing for Demand-Optimised Management |
| **WITS** | Water Industry Telemetry Standards |
| **WoT** | Web of Things |
| **WSSNO** | Water Semantic Sensor Network Ontology |
| **WVCSM** | Water Value Chain Semantic Model |

# 1 INTRODUCTION

## 1.1 PROBLEM STATEMENT

Smart cities and industrial systems, such as power grids, water distribution networks, and district heating networks, are facing increasing pressures towards sustainability, resilience, and service quality [1]–[5]. Artificial intelligence, and ICT in general, have been heralded as the solution to these challenges by offering remote sensing, actuation, and intelligence; using machine processing to assist with the feedback and decision making process [6]–[9]. Recent advances in cybernetics, along with big data technologies, and low cost hardware and communication solutions, have created a wealth of resources for better management of a smart city's system of systems [10]–[13]. However, this increase in ICT penetration within complex systems has led to interoperability chasms between data silos in terms of visibility, syntax, protocols, semantics, security, licensing, and trust [14]–[16]. Even beyond this, the nature of leveraging these resources for business and societal value is not clearly understood, resulting in the idiom "*drowning in data*" [17].

The Internet of Things is attempting to solve the challenge of interoperability [18], [19], and has made significant advances, but there is much work remaining to achieve the level of integration required to truly unlock the value of advanced software in these domains [20]–[23]. Specifically, whilst the Internet of Things is achieving communication-layer interoperability, this does not consider the use of this data at the application layer [24], [25]. This has recently been observed in research, and has led to a new field of studying application-layer interoperability, which has been termed the Semantic Web of Things, owing to its grounding in semantic technologies [23], [26], [27]. Now, significant effort is required to accelerate progress in this field by developing and applying Semantic Web of Things processes and artefacts [28]. This includes the development, adoption and standardization of relevant ontologies, but also knowledge surrounding their lifecycle processes and accompanying software, towards application domain value [29].

Regardless of application-layer interoperability, how can organisations derive value from the 'rising tide' of big data? Ongoing big data and ICT intervention research should keep closely abreast of the growth of the Semantic Web of Things, in order to best capitalize on artificial intelligence [30], optimization [31], [32], simulation [33], and advanced applications in general. It is critical that cyber systems integrate data effectively and efficiently, and user interfaces apply business semantics intelligently, in order to best empower decision makers and organisations [34], [35]. This leads to the challenge of capturing complex semantics, then using them to better support advanced applications, built on the technology of the Internet of Things [36], [37]. Again, this requires significant research around the Semantic Web of Things, and emphasizes the need for a pragmatic mindset of using semantic technologies to provide business value in real-world contexts.

This thesis presents a step towards unlocking the full potential of IoT and 'AI' through semantic web technologies. The work begins from a robust theoretical grounding, before conducting extensive engagement with experts though smart city research projects, and ultimately develops and tests ontologies and software at both the back and front-end. The experimental work iteratively learns from investigations in the (relatively more advanced) energy domain, before exploring the applicability of these in water domain use cases. The investigation finally unifies and generalizes these action and design research iterations through a semantic smart city platform.

## 1.2 CENTRAL CONCEPT DEFINITIONS

In order to fully understand the stated problem, it is pertinent to define some of the key terms and concepts. This section briefly introduces an understanding of smart cities, interoperability, and ontologies, as primers to their discussion through the literature review.

### 1.2.1 WHAT IS A SMART CITY?

The term 'smart city' has many connotations owing to its wide use across the application of digital technologies towards intelligence in urban environment management [38], [39]. The term is arguably an extension of the 'digital city' trend, placing greater emphasis on the *intelligent* use of ICT, hence prioritising the *value* derived from the application of ICT [5], [40]–[42]. Further, pertinent rhetoric typically emphasises the role of citizens and societal benefit. This leads to the notion of ICT as a facilitator, whereby systemic improvement is the goal of smart cities, and digitisation is associated with this due to its potential for assisting in feedback loops and communication. This is the predominant definition adopted within this thesis, where the ‘end goal’ of smart city research is to holistically optimise urban environments across social, technological, environmental, economic, and political contexts.

Whilst local governments are typically the primary stakeholders in ‘core’ smart city projects, the ‘domains’ of smart cities (e.g. smart grids, smart water) are often outside of local government control, yet central to the paradigm. The many conflicting definitions and views on smart cities are considered fully in the literature review, but overall this thesis adopts a broad perspective and typically uses the phrase ‘smart urban systems’ to explicitly include all smart domains.

### 1.2.2 THE MEANING OF INTEROPERABILITY

Interoperability is increasingly a commonly used term, and appears at the surface to simply mean ‘working together’. However, there are many facets of interoperability, and its typical usage does not acknowledge the depths of the term. Just as people ‘working together’ does not mean they are doing so effectively, software components being able to exchange messages does not mean they are doing so effectively [15], [16], [20], [43]. One interpretation of the facets of interoperability is the sum of: i) resource discovery and communicability, ii) resource interface provision, iii) security and privacy, iv) permission & restraints on resource usage (licensing), v) understanding of response syntax, vi) context, meaning and provenance (semantics) of the resources and data, and vii) trust of all the former aspects.

In general usage, ‘interoperability’ often ignores the latter aspects of these requirements, relying instead on ad-hoc and manual interpretation of semantics when building software. By formalising the meaning of data; such as its units, context, and assumed logic, software developers can achieve semantic interoperability between components in a system [44], [45]. This thesis strives towards automated semantic interoperability, to achieve true system integration in a more powerful and scalable manner. By explicitly formalising semantics, machines have the potential to develop interoperability automatically, at runtime [46], [47].

As well as between machines, interoperability should also be sought between machine-human-business interfaces at the highest order. The outputs of applications should integrate seamlessly with the business processes they support

[44]. Currently this involves an extra stage of interpretation by experts, and improving this interface is the topic of human-machine interaction. Again, by capturing rich semantics explicitly, and leveraging them properly at the application layer, the knowledge value chain from sensor to business impact can be far more efficient and effective [48]. Therefore, the definition adopted for interoperability includes a broad scope; including low-order syntactical interoperability, terminological and semantic interoperability, and also business logic and human-machine interoperability.

### 1.2.3 A PRIMER FOR ONTOLOGIES

The word “ontology” has a habit of overwhelming people. The academic leaning of its community has disengaged industry and decision-makers, and definitions like “an explicit specification of a conceptualization” [49], whilst correct, have not helped its plight. But for all the philosophy in their discourse, they are ultimately just a list of statements held to be true [50], [51]. This brief introduction aims to demystify ontologies as a first step in engaging the reader and alluding to their relevance for business and ICT systems.

Ontologies mean lots of different things to different people [52]. This introduction focuses on the models developed for the semantic web and written in the web ontology language (OWL, or OWL2 more recently), although alternative interpretations are discussed in the literature review. Various aspects of ontologies are now introduced; starting with the one stated a few sentences ago:

***Ontologies are a list of statements held to be true***

At the simplest, an ontology is a list of statements, organised and written in a specific way, and assumed to be true by a person or piece of software. Each statement has three parts: a subject at the start (the thing which the statement is about), an object at the end (which can be another thing or a piece of data), and a concept which connects them (called a predicate). As several of the statements will describe the same things, this builds up a network of related concepts which represents a rich knowledge model.

***Ontologies are the data structure of graph databases***

Ontologies provide structure to the data stored in a graph database, where ‘graph’ is the term used for the ‘network’ mentioned previously. In this way, they can be compared to traditional data schemas, where they offer several benefits, as discussed later.

***Ontologies capture a world view***

As a ‘list of statements held to be true’, an ontology captures a specific ‘world view’. As people often disagree on what is true, the highest ambition in building an ontology is to achieve consensus amongst a community that the world view captured is a good representation of a domain.

***Ontologies describe the objects and relationships in a domain***

The subjects and objects in the statements mentioned are the ‘things’ in a domain, which would be represented by nouns in natural speech, such as *building*, *sensor*, or *dog*. The predicates between these are then the relationships between the objects in a domain, such as saying that a *sensor* is *‘deployed at’* a *building*.

***Ontologies are machine readable***

The statements and world view contained in an ontology are understood by machines, and are designed to be easy to build software on top of. This is a key part of how they promote an automated understanding of data meaning and context.

***Ontologies capture expert knowledge for reuse***

As well as just stating the objects and relationships in a domain, an ontology can also contain detailed logic, rules, datatypes, and restrictions, to capture a rich model of expert knowledge. By making this knowledge machine readable, it can be reused by any connected piece of software in a scalable way, rather than relying on an individual person’s expertise and institutional knowledge, which would be lost when they leave an organisation.

***There are ‘domain models’ and ‘instance models’***

Ontology experts distinguish between the part of the model used to describe a domain (e.g. energy grids), and the part of the model used to describe an instance of that domain (e.g. the power transmission system of London). By keeping these parts well defined, the domain part can be kept general enough to reuse in any model of the domain. Technically, it is the domain part which is called an *ontology*, whilst the combination of both of them is called a *knowledge base*.

***Each object is uniquely identified in an ontology***

A key aspect of the semantic web is that each ‘thing’ on the web can be identified uniquely in all contexts, which is achieved through a specific type of identifier. These often resemble familiar web addresses (i.e. http://www.example.com/). This allows a great deal of precision in identifying ‘things’.

***Ontologies separate ‘knowledge logic’ from ‘application logic’***

Ontologies describe the objects and relationships in a domain, but aren’t designed to only suit a single application. They therefore separate the way data is stored, structured, and perceived, from how it is used in for a specific purpose. This separation allows data to be more reusable, and improves the software development process.

***Ontologies are one part of a big picture of ‘semantic technologies’***

Semantic technologies include ontologies, knowledge bases, database technologies, the query language used to interact with them, AI components which reason over knowledge bases to produce new knowledge, and file formats used to store and exchange this data.

### 1.2.4 A NOTE ON ARTIFICIAL INTELLIGENCE

There are many contrasting schools of thought on artificial intelligence, and there is no consensus on the definition of the term [53]. Purists tend to only regard machines worthy of the title ‘artificial intelligence’ if they demonstrate *general* intellect similar or in excess of human intellect. This school of thought started at the birth of AI, where the task of recreating human intelligence was infamously underestimated at the Dartmouth College conference [54], and is also related to the famous Turing test [55]. However, more modern schools of thought tend to apply the term more liberally, through the concept of agency [56], [57].

According to the agent-based school of thought, anything can be considered to display intelligence if it is able to i) sense its environment (i.e. it has an input), ii) apply some form of reasoning to its perception of the environment, and iii) cause an effect on its environment in response to this reasoning (i.e. it has an output). More advanced agent types then carry additional requirements, such that the agent must have specific goals (such as is stated in the belief, desires and intentions model [58]).

The broadening of the definition of AI has stemmed from a focus on the value proposition of AI, where products need not exhibit *general* intelligence in order to provide value to a market. Driverless cars are a topical example of a product which demonstrates intelligence, without resembling the traditional humanoid perception of AI. This thesis adopts a liberal interpretation of the term, and often refers to both 'AI and advanced applications', to explicitly include many technologies such as advanced optimisation and simulation. These would not be considered as AI by some, but are at the core of a broad group of technologies likely to provide great value to smart systems [59], which require a thorough understanding of the semantics of the problem space in order to achieve their intended functionality.

## 1.3 EXAMPLE INDUSTRIAL CHALLENGES AND OPPORTUNITIES

The precedence for Semantic Web of Things research for smart city and industrial applications has been suggested in the problem statement, and its academic grounding and contribution to a research gap is discussed in the background section. Beyond this however, it is pertinent to consider the potential impact and exploitation of the research, within existing business and socio-political contexts. This section therefore proposes a selection of topical industry issues so as to frame the nature of the work conducted within the current landscape of smart city and industrial systems. The key matters briefly discussed are the growth of a distributed energy landscape and the growth and challenges of the smart water market.

### 1.3.1 GROWTH OF DISTRIBUTED ENERGY RESOURCES

Urban energy systems are rapidly changing from centralized systems to the distributed energy systems reported in research. This is partly due to the growth of smart grids (SGs) [60]–[65], distributed energy resources (DERs) [66]–[69], and their accompanying management structures [70]–[73], multi-energy systems (MESs) [74]–[80], and demand side management (DSM) [81]–[84]. This embodies an underlying shift towards sustainability [85], [86] and resilience [87]–[89] through distributed resources, intelligence, and system integration. However, research has only considered these novel concepts within the context of centralized generation [31], [90]. This assumption is becoming unsuitable in preparing for a landscape with many diverse and distributed energy resources, and active consumers. Distributed solutions are currently investigated in isolation from others, or under the assumption of sparse DER penetration [71], [91], but this limits their value proposition. Continuing this assumption will hinder the effective exploitation of renewables and novel concepts, and increase barriers to entering the distributed generation landscape. This calls for a new generation of energy systems which fully embrace a system of systems nature with a tight ICT coupling in a scalable, interoperable and secure framework. Such a cyber-physical landscape would require the interoperation of a vast array of virtual and physical assets.

The importance of interoperability in energy systems is only likely to increase [15], [92]–[94], a large unsolved part of which is semantic heterogeneity between the many software and hardware artefacts. The data formats, terminologies, meanings, and logic used by people and software across disciplines and companies are often incompatible and require ad-hoc mappings to interoperate effectively. IEEE recently emphasised this through a ‘smart grid interoperability guide’, which referred to the challenge of protocol, data format and meaning interoperability [15]. This challenge will become increasingly pertinent as the volume of data and number of software artefacts involved in energy management increases alongside DER penetration, big data growth and the requirement for intelligent management [95]–[100].

In order to overcome the interoperability barrier, secure communication frameworks, and service oriented architectures hold much potential [15]. Another critical piece of the solution though, is the use of a common vocabulary and data model. This mitigates the effort required for software artefacts to communicate effectively, helps in integrating legacy systems, and promotes security and performance [15], [101]. These common models must standardize descriptions of the concepts and relationships in the domain, as well as terminology, logic, and data formats.

As mentioned, the IEEE 2030 standard is a seminal development in the energy interoperability discourse [15]; it provides normative guidelines and a set of term definitions and descriptions through a ‘Smart Grid Interoperability Reference Model’. This provides essential groundwork by facilitating a common human understanding and advising on best practice. The guide also highlights the role of ontological models in ensuring a shared meaning of data, hence making it more valuable, as well as providing inference and rule-based functionality.

Given the precedent from industry, growing academic interest, and the ongoing evolution of power systems, it is critical that energy system interoperability is addressed through research, including a focus on semantic modelling. Whilst ontologies are heralded as critical, their development, adoption and standardisation for energy systems is still “embryonic” [102]. Research towards a highly expressive and flexible urban energy model is critical, as well as investigation into how best to leverage such a model to improve energy system operation.

### 1.3.2 SMART WATER: GROWTH AND CHALLENGES

Water shortage has been named as the 3rd biggest risk factor for the next decade by the Global Economic Forum as of 2016 [103]. Even in areas of water abundance, the water sector faces mounting challenges. Utility companies are facing increasing regulation [104], and demand is growing and becoming more concentrated, due to increasing populations and urbanisation. In the UK specifically, business deregulation in 2017 poses imminent market disruption [105]. Managing capital expenditure is a critical issue for utilities, as well as minimising energy consumption, avoiding regulatory fines, and maintaining a positive customer perception [106].

These challenges have led to growing pressures to operate, maintain, and invest in water networks as intelligently as possible. ICT is being explored as a means to reduce OPEX and lengthen the lifetime of aging infrastructure, to mitigate the need for additional infrastructure, also reducing CAPEX [107], [108]. This penetration of ICT has parallels with smart energy grids, and has been termed ‘smart water’. New research and products are emerging across the technology stack to empower decision makers and protect our water supplies. Smart metering is one of the most publicised aspects of this [109], and intelligently managing network pressures to reduce non-revenue water and pipe bursts is another promising example [108].

Smart water is a growing market [110] and offers significant potential value to stakeholders in water systems. However, interoperability has again been consistently flagged as a roadblock against unlocking this promise [107], [111]–[113], similar to in the energy sector [15] as mentioned previously. A number of steps have been taken towards overcoming the water interoperability challenge, including the WITS protocol [114], but progress is slow, behind the energy sector, and only focusing on low-level interoperability. Higher order interoperability has only recently gained interest, but has since been stated as the most important hurdle to overcome [111], and as such is the focus of this thesis. By learning lessons from the energy sector, and elsewhere, and addressing this issue in an open, scalable, and robust way, smart water will be well positioned to continue its growth and deliver its full value potential across stakeholders.

## 1.4 RESEARCH OBJECTIVES AND SCOPE

Based on the introduced need for research, the current work aims to take a step towards filling this gap. This is formalised through an overarching hypothesis, and its decomposition into specific research questions.

The hypothesis to be tested is:

*A Semantic Web of Things approach to technology interventions can deliver value to smart city stakeholders by better leveraging IoT and AI synergistically to provide better decision support.*

In pursuit of a knowledge contribution towards evaluating this hypothesis, within the stated scope, the work will aim to answer the following research questions:

1. What are the theoretical underpinnings of ICT knowledge gaps in smart cities, including the challenges, impact scenarios and scope for step changes?
2. How can a semantic web of things approach integrate IoT and advanced smart city applications?
3. What value does a semantic web of things approach offer technology providers and decision makers in smart city systems?

4. Can these learnings and artefacts be generalised to support further work across smart city domains and semantic web of things research?

## 1.5 THESIS OVERVIEW

The main body of the thesis now continues in a style which is less common within engineering disciplines. Due to the approach's emphasis on iterative learning and a mixture of participatory action research and design science, the importance of the work to the primary contribution increases towards the final iteration, which is presented in Section 4.4. This is summarised in the hierarchy presented in Figure 1.

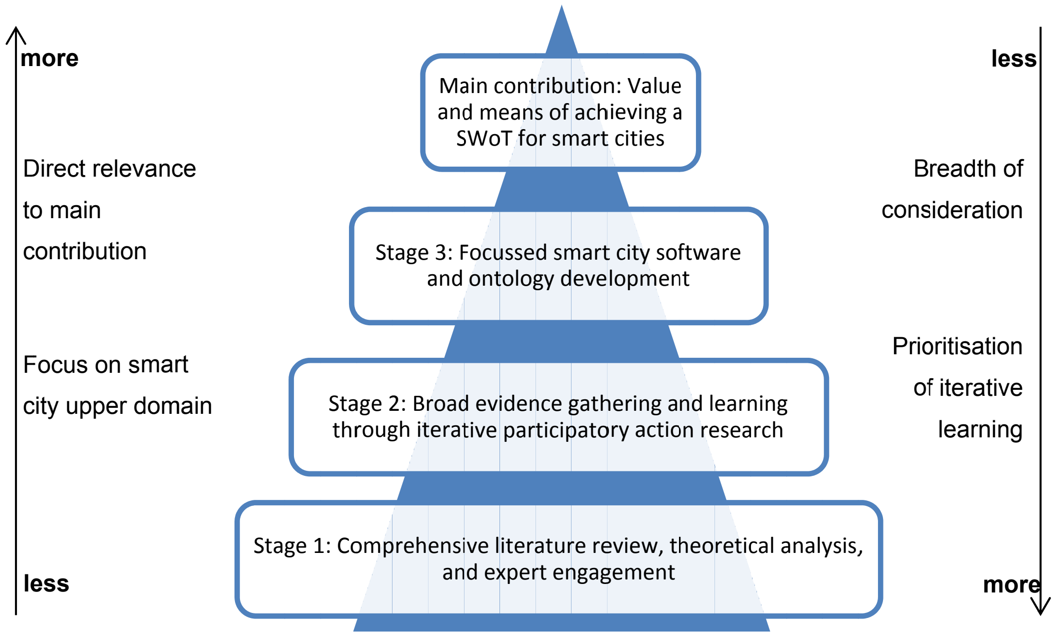


**Figure 1: Heirarchy of the relevance of the work conducted to the central contribution of the investigation**

The following chapter presents a review of the literature and argues towards the gap indicated previously. This begins by discussing the value proposition of smart cities and smart systems in general, in terms of their drivers for change, as well as various pertinent conceptualisations and models, before examining case studies. Next, the role of the Internet of Things is examined as a canvas on which system theories can be effectively leveraged. This considers interoperability standards in depth, discusses the latest research, and identifies challenges in the field. Various perspectives are then offered of semantic technologies from the literature. Finally, the manifestation of these concepts in industry is analysed within smart energy and water grids, contextualised within the latest developments in the energy and water sectors.

The 3rd chapter presents the systematic methodology followed in conducting the investigation. This begins by building a robust theoretical grounding in the nature of truth and its investigation through scientific methods. The chapter then transitions to the manifestation of the adopted research philosophy through an overarching approach, before providing a pragmatic overview of the work conducted. The chapter then proceeds to describe in detail the processes undertaken through the investigation, as a sequence of 3 parts: theoretical study, iterative participatory action research, and independent design research.

The 4th chapter presents the results and outputs of the investigation, grouped by the stage of the investigation. Firstly, this presents the outputs of the theoretical study undertaken, which includes high level scoping and impact scenarios, as well as a high-level conceptual framework which guided the rest of the investigation. The software artefacts and ontologies contributed to in the 2nd stage are then described and illustrated, and quantitative data from their testing are also included. Finally, the semantic smart city platform developed in the 3rd stage of the investigation is presented, again in terms of its software artefacts, ontological modelling, and performance testing.

Chapter 5 then discusses the results and outputs produced. This begins by analysing the individual projects engaged with and arguing towards the main research findings. The chapter then considers the overall study against the literature, from the perspectives of IoT, smart city, energy and water domain research. The chapter next outlines the relevance of the contribution to current practices in academia, local authorities, and industry.

Chapter 6 concludes the thesis with a final discussion of the work conducted and results obtained, from the perspective of the initial research gap identified, framed through the hypothesis and research questions. Firstly, the main research findings are summarised, before this is unified into a brief section outlining the key contributions to the body of knowledge. Finally, the chapter discusses the limitations of the investigation and proposes some potentially valuable future work.

The list of references is included after the conclusion chapter, followed finally by the appendices. Information about the motivation of the author in pursuing this work, and of the target audience, can be found in the appendices.

# 2 LITERATURE REVIEW

This section identifies the literature gaps which resulted in the research questions posed previously, and contextualises these within the broader research field. First, smart cities are described as a system of systems with a value proposition that currently isn't being realised due to a lack of shared semantic referential and powerful knowledge exchange, which has prohibited the use of IoT within advanced applications. The IoT field is then explored in depth, and integrative application layer interoperability research is argued for, so as to move beyond the current IoT capabilities of low-level data exchange towards higher order business intelligence. The role of semantic technologies in delivering the potential of IoT and smart city concepts is then discussed in depth, and are related to the surrounding fields of AI, business intelligence, big data, and information systems. Existing smart city ontologies are then considered, before using smart energy systems as an example domain where evidence of the previous arguments can be observed. Finally, the emerging importance of semantic technologies in the water domain is discussed to highlight the growing nature of this field across smart domains before concluding remarks.

## 2.1 THE VALUE PROPOSITION OF SMART SYSTEMS

The smart city paradigm has gained popularity recently as a broad concept, where the term 'smart' is related to the terms 'digital', 'eco', 'ubiquitous', and 'future', in this context. Much technology focused research is labelled with 'smart city', and mostly this aims to improve the operation of a city system, and possibly encourage citizen engagement. This emphasises that the 'smart' label broadly refers to 'value delivery through intelligence', distinct from the similar terms. The main stakeholders in smart cities are citizens, local authorities, utility companies, and technology suppliers. The main purported benefits of smart cities are: personalised services, better access to information, efficient management, better decision making, better service provision, and citizen engagement [7]. Some of the key intended outcomes of making cities smarter are the promotion of high-quality, robust, and secure service delivery across systems, whilst simultaneously reducing the city's environmental impact and increasing the perceived benefits of citizenship [42], [95], all whilst improving economic viability. A key aspect of this is the management of resource flows;

including the key resources of energy [69] (electricity, heat, and raw fuels) and water.

This section begins by discussing the etymology and various interpretations of the term ‘smart city’, before examining the drivers for change and various high-level models of smart cities. A number of case study cities are then discussed, before exploring the system of systems nature of urban environments. Finally, an argument is built towards the need for knowledge exchange between agents in order to manage this complexity successfully, especially whilst leveraging ICT.

### 2.1.1 ETYMOLOGY AND HIGH-LEVEL PERSPECTIVE OF SMART CITY CONCEPTS

The label of ‘smart’ has increasingly been applied to research regarding built environment systems, as shown in Figure 1. The term is associated with the goals of sustainability, adaptability, efficiency, prosperity, and citizen satisfaction, amongst many others. However, there is no agreed meaning [1], [7], [115], and the term is often conflated with several related terms, such as ‘green city’, ‘future city’, ‘eco city’, and especially ‘digital city’. The origins of smart cities can be tracked to the ‘digital city’ paradigm [5], which Figure 2 shows declined in general popularity circa 2004-2008, whilst the ‘smart city’ term boomed circa 2014 in web searches. However, the evolution arguably began in the 1990s [116], although ‘smart’ can be considered a rebranding of cybernetics, which originated far earlier [117]. It is also interesting in Figure 2 that whilst ‘smart city’ is a global term, the other terms are more localised in popularity, mandating a consideration of these alternative terms, which are used *instead* of ‘smart city’ in some countries.

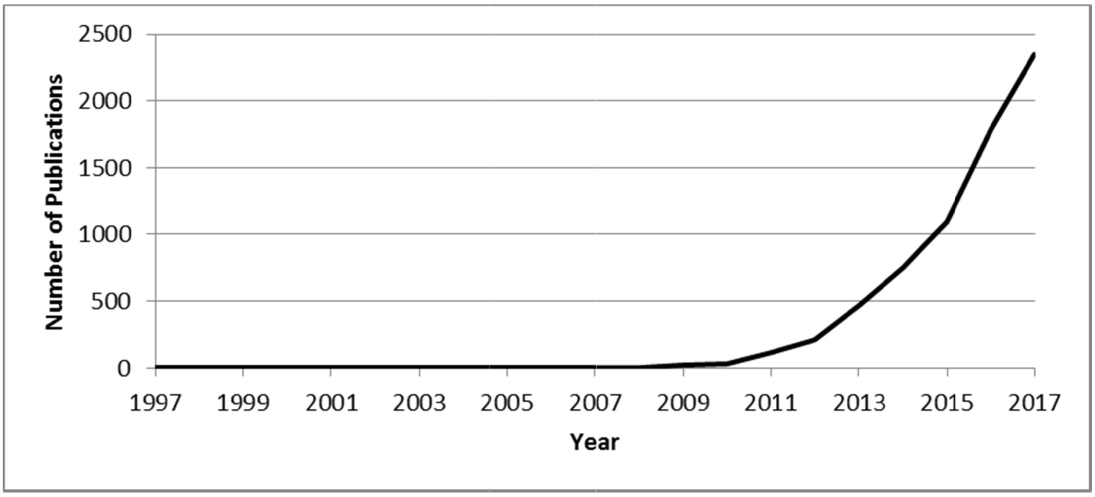


**Figure 2: Results of Scopus search for TITLE-ABS-KEY("smart city"), 2017 estimated by extrapolation of year-to-date**

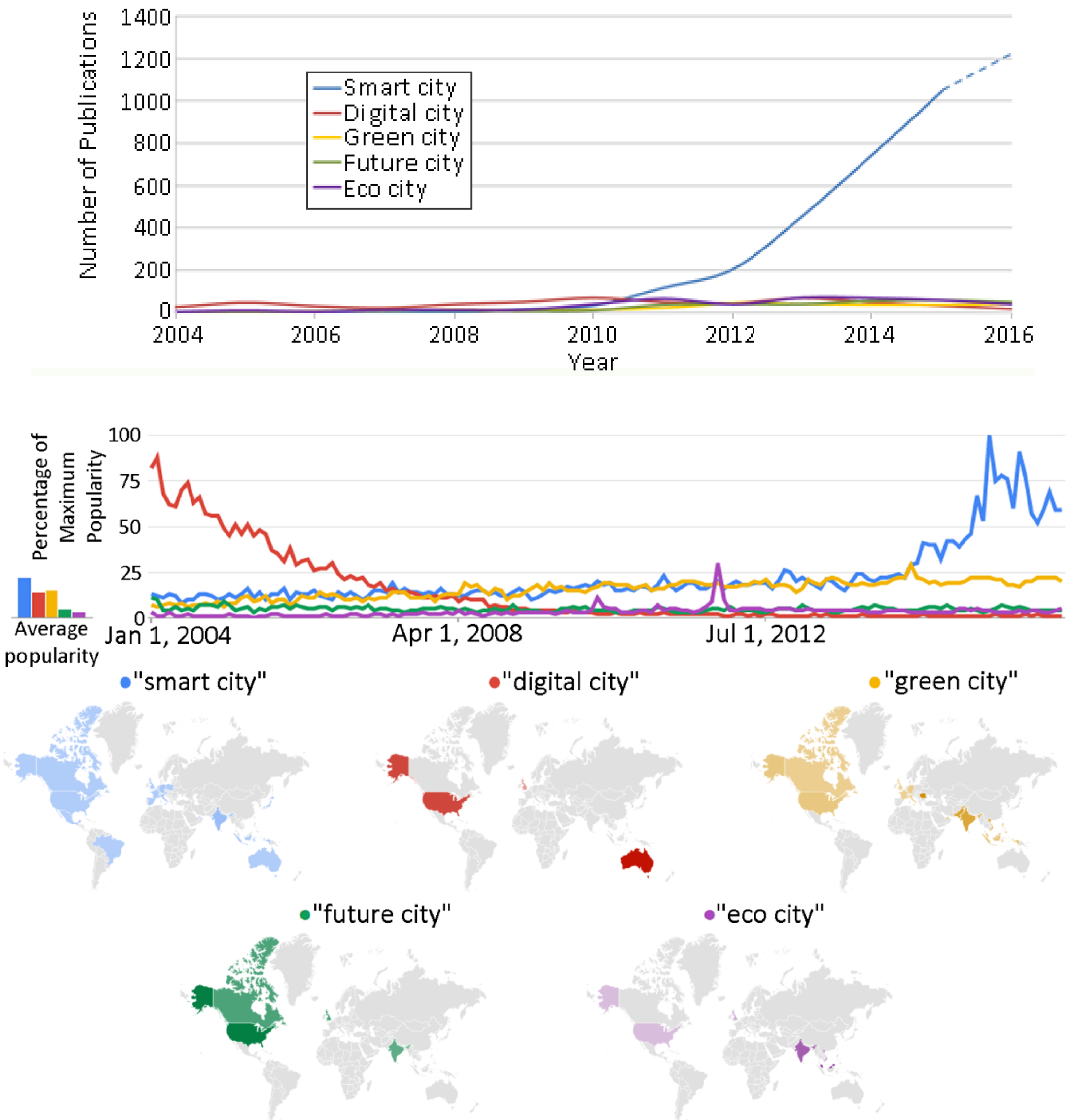

**Figure 3: Growth of the 'smart city' term relative to similar terms in Scopus (top) and Google [118] (bottom)**

Central to smart cities is the use of data and analytics in multi-objective interventions across technological, environmental, productivity, quality of life, social, and infrastructure performance indicators [119]. However, confusion arises from the term's similarity to the 'digital city' trend. Su [40] states that whilst a digital city encompasses the city's sensing and GIS infrastructure, a smart city builds on this to deliver applications and improve system performance. This promotes the view that 'smart city' research should focus on the value proposition of intelligent management, made possible *through* ICT, which agrees with Yeh [41] and Batty et al. [42].

### 2.1.2 THE SMART BUILT ENVIRONMENT

A smart built environment integrates intelligent, informed, and evidence-driven decision making across the lifecycle of a building or other built asset, from initial concept, through design, construction, operation and end-of-life processes. This is a multi-criteria decision process, where technology should aim to provide the most valuable insight possible to decision makers regarding service delivery, capital and total expenditure, energy performance, and holistic considerations such as social and environmental impact. This vision has been partially captured by the recent industry trend of Building Information Modelling; a knowledge-driven and digitised set of AEC processes and technologies which also has value in asset management (AM) and facilities management (FM).

From a knowledge modelling perspective, BIM represents the evolution of CAD to include not only geometric and aesthetic properties, but also semantic information regarding the entities, processes and actors which the 3D forms represent. BIM initially aimed to facilitate knowledge exchange from design to construction, but has been increasingly researched at operational phases of assets. This has included the use of the Industry Foundation Classes (IFC), the *de facto* data model of BIM. However, much work remains to integrate these concepts and standards with the holistic smart building vision. From a process-oriented view, the prioritisation of operational concerns at design stage and the role of the asset within the wider smart community should be researched and standardised further, to minimise total expenditure, environmental footprint, and embody smart principles generally. From a technological perspective, the data structures and software of BIM should be researched alongside the IoT and AI fields in an integrative manner to leverage these valuable paradigms maximally within buildings.

The operational lifecycle phase will be focused on here, where technological interventions are primarily made through a building management system (BMS) or building automation system (BAS). For example, research has studied the role of sensors, local PID controllers, building automation, artificial intelligence, rule based control and fuzzy logic [120]. Research also includes evolutionary based optimisation, model predictive control and multi-agent systems control at various scales between local and supervisory [121]–[124]. However, this research is typically performed *in vitro*, and the integration of IoT with advanced applications is not observed in commercial products. This section forms the argument that this is due to a lack of integrative research between the relevant aspects of the IS, IT, and AI fields, where research on semantic technologies could provide a bridge to overcome this obstacle.

### 2.1.3 SMART CITY DRIVERS FOR CHANGE

Following from an understanding of the origins and high-level themes of smart cities, it is pertinent to consider the driving forces behind the trend, before reviewing more detailed conceptualisations of smart cities, and case studies of relevant cities, in subsequent sections.

The main stakeholders behind the emergence of smart cities are arguably governmental bodies, from supranational to local scale, and large technology companies, although city leaders and citizens also play key roles. At the international level, Europe has been referred to as the global centre for smart city research [1], [6], although [125] notes that whilst a driving force, European Cities tend to emphasise soft aspects and human capital. Outside of Europe, South Korea has strongly pursued the related ‘ubiquitous city’ agenda [126], [127], China and USA are also at the forefront, with other notable examples including Singapore [128]. The UK can also be argued itself to be leading the field with prominent example cities such as London, Edinburgh and Glasgow [129] as well as the recent BSI smart city ‘suite of standards’ [7], [16], [130]–[132] which recommend best practice across the smart city field and emphasise the importance of an interoperable data framework.

Pertinent drivers for change include social, economic and political pressures, and the need to be attractive places to live to retain the ‘best’ citizens [133]. Further, national agendas to ‘lead the way’, place political pressures on cities to become smart so as to improve the country’s perceived level of development and again to attract talent and business. Finally, business motivations behind making cities ‘smart’ must be acknowledged. Many companies have driven the smart city agenda, from large multinationals to start-up companies which have emerged specifically to benefit from this growth market valued at $1.5 trillion [134]. The most commonly referred to market leaders in terms of innovation and perceived market share are IBM [135], who arguably began the movement through their smarter planet initiative, as well as Intel, Siemens, Microsoft, Cisco, and Arup, (in no particular order), amongst others.

Within this political and business context it is essential to make efforts towards a neutral consideration of the field to form valid scientific sentiment; as making a city smart can occur in pursuit of furthering a variety of agendas. It is also pertinent then to note the bottom-up drivers of smart cities. Specifically, whilst public bodies and large ICT companies often wish to imbue smartness in cities, other companies, organisations, and individuals often wish to promote aspects of the smart city paradigm for their own reasons. This includes industrial systems, whereby smart grids and smart water networks, for example, promise significant benefits to the relevant stakeholders, and to empower customers with choice and an active role in the sector both economically and technologically.

Finally, the role of the individual must not be understated, whereby improving the ‘smartness’ of a city should involve a symbiotic relationship between improving the quality of life of citizens and engagement of citizens in city management. For example, local hackathons to create apps based on local open data can serve to engage citizens and deliver genuine value to them through applications made ‘by the people, for the people’.

### 2.1.4 CONCEPTUALISATIONS OF SMART CITIES

Within the smart city field there are a multitude of high-level summaries and 'clarifications' of the composition of smart cities. These indicate the themes surrounding smart cities, but they mainly serve to indicate their author's perspective. One common theme in describing smart cities is the use of a layered model to visualise and logically categorise the various entities, with a focus on the use of knowledge and ICT systems to deliver business services or applications [8], [40], [126], [136]–[141].

Layered models of smart cities generally describe the flow of knowledge from sensors through an ICT system, to ultimately interface with users. [139] presents an example model, where data is collected from sensor nodes and processed through storage and security systems before arriving at analytics engines and interfacing with applications. [138] presents a variation on this, which includes green initiatives and an 'innovation layer'. Simpler layered models include those of Yovanof and Hazapis [140], and Su et al. [40]. Su et al. proposed a model with the layers of *perception*, *network*, and *application.*

These models are arguably just variations of the seminal work by Harrison [8] on IBM's smart planet initiative. He states that smart cities must be "interconnected, integrated and intelligent". 'Interconnected' refers to sensor networks; either physical or virtual [8], where virtual sensors measure virtual quantities such as an occupant comfort index or the number of 'tweets' from a location. 'Integrated' then refers to simultaneous analysis of heterogeneous data sources, and 'intelligent' refers to the application of analytics to determine the best course of action. As a novel interpretation of the paradigm though, the model of Komninos [142] emphasises the knowledge and innovation systems within smart city 'environments', which include e-markets, and e-technologies. This breadth of thinking around smart cities, is represented well by the BSI 'integrated operating model' for smart cities, within their smart city framework publicly available specification (PAS) [130].

The discussed views agree generally with the layered model presented in the review of the UK future cities demonstrator proposals conducted by the Technology Strategy Board (TSB) and Arup [143], which shows most cities including a web-based virtual platform on top of an ICT infrastructure. However, this diagram implies a broader consideration of the underlying city context and also relates to several models of cities as system of systems. This concept of smart cities as an ICT-facilitated manifestation of system theories is now discussed in the following section.

### 2.1.5 REVIEW OF SMART CITIES AROUND THE WORLD

Grounded in the view of cities as an interconnected system of systems, this section examines a number of cities which have been described as 'smart cities', as an overview of the global state of the art in practice. The cities were chosen loosely based on their prevalence in the discourse. The section starts with UK cities Glasgow and London, before looking to another European city of Barcelona, and then further afield to Singapore and New York.

#### 2.1.5.1 GLASGOW

Following a £24m award in 2013 [144], Glasgow became the UK's demonstrator of smart city innovation, with the aim of continuing its transformation to a vibrant, diverse and modern city, beyond its current health and social challenges [143].

With a focus on leveraging data and citizen engagement, Glasgow developed an operations centre, open data hub and a city dashboard to allow citizens to interact with the city and develop their own applications. These were collectively termed a 'City Technology Platform', which also encompassed GIS elements, and data hub APIs and platform services [145]. The operations centre coordinated public services related to safety and traffic [146]. The data hub offers a real time dashboard, as well as exposing and describing circa 400 data sets [147], which are mostly infrequently updated CSV files. These consist mostly of public sector data such as energy consumption of public buildings and public car park occupancies. The main use of these datasets appears to be a somewhat primitive city dashboard and a number of interactive maps.

Data management within the future city Glasgow project appears to be ad hoc. The data catalogue doesn't address the semantic context of its data, although there are plans to allow the upload of linked data in the future [148]. This currently limits the value of the data hub, as finding datasets is limited to simple search methods, and the retrieved data is heterogeneous. This makes it difficult to know what data is available and requires significant effort to homogenise data at the application layer. In the initial Technology Strategy Board Future Cities Demonstrator proposal, Glasgow indicated it would utilise an 'urban ontology of asset management' [149] and a 2014 presentation indicates this is still the case [145]. From a review of the relevant webpages, it is likely this has manifested in an asset map for Glasgow [150], if this is the extent of its utilisation it fails to leverage the majority of the benefits which semantics offer; utilising this ontology to contextualise the datasets would be a significant improvement.

#### 2.1.5.2 LONDON

London is expecting to increase in population by 1 million citizens between 2011 and 2021, and to continue this acceleration towards 2030 [151]. London is considered to be paving the way in smart city technologies [152], in part due to the efforts of 'Transport for London' [153], [154] such as its oyster card scheme and advanced bus network, as well as London's effort on district heating systems [155]. London appears committed to following a smart city agenda with data, technology and people at the core [151], [156], [157] in order to tackle challenges such as traffic, unemployment, energy supply, waste management and pollution [151].

A 'smart London' plan was established following an award of £3 million as a runner up to Glasgow in the Future Cities Demonstrator competition. This plan describes its 7 pillars as i) citizens at the core, ii) open data, iii) use of human capital (aka the creative class/innovation centric), iv) links between projects, v) innovative projects, vi) integrating City Hall, and vii) overall 'smartness' [151]. The plan also emphasises system of system approach, defining 'smartness' as a measure of how the overall system functions as a result of its subsystems.

It is interesting to observe a similar effort to Glasgow and Barcelona in producing a centralised repository of data, here referred to as the 'London Datastore' [158]. This is similar to the Glasgow Data hub in that datasets are accessed via a simple search engine, and are tagged with topics, formats and publisher. This is useful as a first step in order to make the data available, but could be improved through semantic search or an object-oriented API. This approach of providing data as separate entities has been described as a 'bottom up' approach similar to those in Bristol and Leeds [159], but does not recognise the importance of providing a framework to contextualise the data. By unifying the data landscape in a more comprehensive way, barriers to its use could be drastically reduced.

#### 2.1.5.3 BARCELONA

Recently, Barcelona has been ranked as the top overall smart city in the world by Forbes [160] and Juniper Research [161], one of the top 10 by IMechE [152], the best example of intelligent city infrastructure by CCLA [162], and the 3rd most climate-resilient city by Fast Company [163]. Barcelona prides itself on being a centre for innovation [163]–[165], and has brought together a number of smart city projects into a unified goal-oriented strategy [166]. Barcelona's smart initiatives have been grouped into "Smart Governance, Smart Economy, Smart Living and Smart People" [167]. This includes all of the typical urban domains, although addressing research and tourism within a smart city strategy is novel [168].

Within Barcelona's range of now unified projects, it aims to provide Wi-Fi to all citizens, more charging points for electric vehicles, hireable bicycles, government eServices, and opportunities to develop apps through open data and hackathons. Barcelona's 'City OS' project provides an interoperability and intelligence layer [169], which forms one part of BCN's 'smart city platform' alongside Sentilo (which aims to capture data) and Applications, in a familiar layered arrangement. The City OS 'city semantics' module [170] is very interesting, and utilises an ontology to "organize data", and decouple raw data from the applications which use it. There is also significant interest in urban data & resource integration through ontologies & big data science from the Barcelona Supercomputing centre [171].

Despite its accolades, Barcelona is not without its critics. Moskvitch recently argued that Barcelona's smart interventions appear "bitty and piecemeal"; more like a showcase of potential than "part of the fabric of city life" [172]. This resonates some of the criticisms of smart cities as a whole; that a bias towards ICT penetration does not necessarily improve the underlying city or quality of life for citizens, and mandates technical literacy to achieve any benefits. It is imperative therefore, if the 'smart city' trend is to deliver value to citizens and stakeholders, that those people and organisations are put first in planning smart interventions, rather than looking for opportunities to deploy ICT in a city.

#### 2.1.5.4 SINGAPORE

Another city which has frequented the top of ‘smartest city’ lists is Singapore, which has actually launched a ‘smart nation’ initiative in 2014 [173], and has a long history of striving to be smart [128]. The initiative has very recently announced a new level of sensing deployment with the aim of monitoring public spaces and citizens for more comprehensive real-time performance monitoring and predictive analytics [174]. This has caused significant privacy concerns, although the political climate is significantly different in Singapore to countries where public freedoms are given the highest merit, and in that Singaporean state-owned companies are far more pervasive than in western societies.

Smart Nation Singapore is split into 4 initiatives; Health, Living, Mobility, and Services [173]. The health initiative includes a personal health portal which integrates public health data with data from wearables and nutrition apps. The mobility initiative includes the ability to book a self-driving shuttle from a smartphone app, and ‘on-demand’ busses which adapt their routes to suit community demand. Research has used Singaporean public transport data to understand mobility patterns [175], and used traffic and parking data to assist drivers [176]. Singapore also offers a range of open data platforms, including an API based on the specifications of the Comprehensive Knowledge Archive Network (CKAN), and a geospatial API, although these follow the pattern of other ‘data hubs’ in not providing a shared schema or framework for the data or semantics, and so represents a step change rather than the final state.

A recent review of Singapore’s Smart Nation program is offered by Chia [177], who emphasised the IoT and cyber-physical aspects of the program, and highlighted that the main challenge is likely to be societal rather than technological. However, one key technological barrier noted is “data heterogeneity”, and the need to support many different applications and devices. The main societal issue identified is cultural inertia in breaking down information silos and business boundaries, which is a common concern in smart city rhetoric.

#### 2.1.5.5 NEW YORK

New York was recently ranked by the Institute of Higher Business Studies of the University of Navarra as top city in the world according to their 'cities in motion' index, which measures sustainability and quality of life, and was ranked as the smartest US city, ahead of San Francisco and Boston [178]. This is in spite of New York's low 'social cohesion'. However, the publicity surrounding smart cities is significantly lower in the USA than Europe, despite the US government launching a smart cities initiative in September 2015 and adding more than USD$ 80 million of investment in September 2016 [179]. The White House has stated that the key areas are climate, transportation, public safety, and transforming city services, with an emphasis on smartness through community-driven collaboration. New York is also leading an initiative of 21 cities to ensure responsible and equitable deployment of smart city technologies.

A recent report to the president from the President's Council of Advisors on Science and Technology (PCAST) discussed some of the New York based smart interventions [59]. These included the New York Fire Department using data mining to predict potential fires, based on 7500 factors from 17 city agencies. New York is also involved in 'vision zero', which aims to use data to reduce the danger of "automobile-based transportation systems". New York is also using data about hospital admissions to determine areas of high asthma incidence for air quality monitoring purposes. The City of New York has collaborated with Cisco to offer 'City24/7', a tool which integrates data from various hyper-local sources for display on smart screens, located at previously unused street furniture, or by Wi-Fi on nearby mobile devices [180].

### 2.1.6 HOW SMART IS TOO SMART?

As indicated in the previous subsection, smart cities are not without their criticisms [181], [182], and as smart cities couple the concepts of *digital* and *purposeful*, the criticisms can be split into those against digitisation, and those around the purpose of smart cities, although the former are more prevalent. These criticisms are now introduced and their relevance to this thesis discussed.

Firstly, the smart city trend has been criticised for its bias towards ICT [117], [183], as this may ignore other, better avenues. For example, planting trees along the pavement may offset carbon footprints and improve communities [184], without any need for ICT. Next, ICT penetration is often linked with privacy and security issues, as more ICT typically means a larger 'surface of attack' [185], [186]. ICT penetration also requires that citizens be technically literate in order to benefit from the intervention, which requires special consideration in order to benefit the elderly and those less able to gain ICT skills [187], [188]. There is also great concern amongst some that increasing amounts of ICT, and smart metering specifically, leads to dangers related to electromagnetic radiation [189], [190], although scientific discourse largely discredits these claims [191]–[193].

Regarding the purpose of smart cities, some people criticise that smart cities tend to over emphasise the economic benefits of interventions and strategies, ignoring the societal and environmental impact [183]. Further, smart cities have been criticised for being driven by large technology companies, rather than community problems [183]. Recent rhetoric has emphasised the central role of citizens in smart cities, perhaps in response to this concern. Significant concerns have also been raised regarding privacy in relation to surveillance, especially with regards to machine intelligence [194] and big data [195]–[197]. Finally, smart cities have been criticised for a lack of clear purpose, as city 'smartness' cannot then be properly evaluated or compared [38].

The criticisms raised against smart cities do not invalidate the precedent set for striving for 'smartness', nor the significant progress made, but they do offer invaluable context for the hypothesis and the work to be conducted. Specifically, it is clear that the broader context must be considered around the information system interventions undertaken, including especially the privacy and security aspects, as well as the ethical implications. It is clear that a human-centric view of IS goals should be adopted, which should manifest as citizen-centric goals wherever possible and relevant. This will mandate a rigorous consideration of the value of the knowledge produced in terms of real-world problem solutions spaces, as opposed to simply identifying where the knowledge may be applied, as this may ignore better non-ICT solutions to those problems. By using criticisms of the smart city movement to guide the conducted work, the vast predicted growth of IoT, smart city, and related markets and technological capability may be affected for the better for secure, sustainable, and citizen centric improvements.

### 2.1.7 SMART CITIES AS A COMPLEX SYSTEM OF SYSTEMS

In order to conduct research into intelligent urban management, it is pertinent to establish a grounding in the nature of the intelligence to be applied to cities, prior to the added complexity of implementing this via ICT interventions. One key field is systems theory, which has been widely considered within 'smart system' literature [198]. Systems theory attempts to understand the behaviour of an entity composed of a number of bounded parts, and stems from the more natural observation of physical systems [199]. This abstracts cities into systemic conceptualisations of interworking parts, through necessary assumptions and simplifications. By managing the complexity of cities in a methodical and robust manner, the behaviour of cities can be better understood, analysed, and predicted.

#### 2.1.7.1 INTRODUCTION TO SYSTEM THEORIES

Von Bertalanffy proposed the original 'general systems theory' [199], from which systems theory has been developed to suit desired applications, for example system dynamics theory focuses on feedback interactions between system components [200]. System dynamics uses causal loops and feedback mechanisms to model complex systems, and have been used in DSTs for building management [200]. Also, Park et al. used system dynamics modelling to investigate design-build strategies in Korea [201], stating that system dynamics allows the application of control theory to industrial systems. Neighbouring the field of systems theory, complexity theory focuses on measuring and reducing the uncertainty in a system [202] and is derived from axiomatic design.

Leveraging an understanding of systemic behaviours can be highly beneficial in promoting sustainable and resilient solutions. Given this, resilience engineering has received much attention in recent years. Many authors have investigated resilience in systems, and frameworks for resilience thinking [203]–[206]. Fiksel defines resilience as a property of systems which exhibit certain characteristics, of which sustainability is critical [204]. This somewhat vague definition is clarified by Wright et al. who define it as the ability to absorb disturbance and consequently reorganise

[206]. Hilton et al. extend this to state that resilience is not the purpose of a system but rather a characteristic which requires certain functions and features [203], concurring with Fiksel. Urken et al. add further complexity by distinguishing between robustness and resilience [205]; robustness is the attribute of functioning normally despite disturbance whereas resilience integrates both robustness and sustainability, which again concurs with Fiksel.

Holonic system theory is based on the concept of a dynamic hierarchy of holons, where each holon represents an autonomous and self-contained system, but can contain or be contained within other holons, within a flexible hierarchy. In this way the holonic approach is a hybrid between the distributed approach where autonomous subsystems adapt within a static framework, and the centralized approach where subsystem behaviour is prescribed by a supervisory controller. The concept aims to balance the objectives of individual systems and the overall system of systems, and originates from the Greek words of “holos” and “on”, meaning “whole” and “part” respectively [207]. Another recurring concept in the literature is the view of cities as ‘emergent’ systems, whereby the nature of the ‘whole’ emerges from the nature and interactions of its ‘parts’ [208], which featured in industrial systems research before the turn of the century [209].

#### 2.1.7.2 SMART CITY APPLICATIONS OF SYSTEM THEORIES

Many recent works have applied system theories within smart city research. For example in energy systems [210], the authors stated that using system engineering practices can reduce risk in projects, and improve return on investment. Dodgson and Gann [3] also advocate the integration of systems of systems. The view of cities as complex systems of systems is also expressed by Javidroozi et al. [211], who draw parallels between city system integration enterprise system integration. The authors use these parallels to propose a business process centric model. Also, they highlight the difference between *interconnected* and *integrated*, where that latter goes beyond exchanging messages, to truly integrate components at a process and conceptual level.

A city can be thought to contain not only technical systems such as an electricity grid but also social, economic, environmental and political contexts [5]. There is a considerable lack of research which holistically considers cities across multiple systems (resource flow, crime, transport) and contexts [32]. One example of a multi-contextual system of systems model of smart cities is presented in [136], which shows how various issues impact across city systems. As well as considering systems across STEEP contexts, it is also pertinent to consider cities as interworking industries [212], and also as interconnected systems of physical infrastructure [213].

### 2.1.8 LEVERAGING SYSTEM THEORIES THROUGH DECISION SUPPORT AND POWERFUL INTEROPERABILITY

The empowerment of decision makers through DSTs [34] is critical to smart cities. Such a tool must deliver insights which are relevant, timely, accurate, and accessible, whilst prioritising user experience. Based on this, current tools such as dashboards and reports are lacking [214]. Dashboards tend to display data, trends, and possibly highlighted problem areas [215], and are generally either geography-centric [216], or object oriented [217], although the convergence of these would be valuable. Further, they typically deliver low-order knowledge such as raw data, performance graphs, or infographics. These only hold limited direct value: the user must apply expert knowledge and analysis to derive higher-order knowledge and business value. This process is unwieldy or impossible in the scenario of optimizing across complex, multi-discipline systems of systems.

Higher order knowledge must be presented to experts, such as contextualised problems, suggested actions, and their implications, through multi-criteria analyses [218], based on real-time data from many sensors across enterprise systems. These next generation decision support systems require the use of advanced applications [219]. Optimization, simulation and artificial intelligence are mature examples [221], but are underused in this domain. Multi-agent systems are also highly relevant, as they are able to balance local and global objectives through emergence, respect information privacy, and be highly scalable [222]. There is a clear gap of intuitive higher-order knowledge delivery tools in the domain. Also, deploying these applications across *in vivo* sensor networks and enterprise systems requires seamless and powerful interoperability. In recent times this has been proposed through data warehousing, but this doesn't address several key issues, so more modern approaches have arisen, such as scalable NoSQL systems [223] and semantic data lakes [224]–[226].

In order to systematically leverage ICT optimally, it is critical to adopt appropriate domain conceptualisations in building decision support software. This requires a careful consideration of the role of technology alongside business processes and physical infrastructure. Whilst this is already a part of software requirements engineering [227], further progress is required to truly integrate heterogeneous technological and physical subsystems with social and business systems. For example, Hefnawy et al. [228] propose a lifecycle based smart city framework for service integration, and highlight the importance to integrate heterogeneous resources. Also, the DAREED project has developed a knowledge-based decision support system which uses multi-objective optimisation and prioritises integration with business processes across stakeholders [229]. This integration of perspectives is also the emphasis of research by Liu et al. [230], following their extensive review into decision support. They also indicate that common models and aligned semantics are essential to achieve a multi-faceted integration including services and processes, and also highlight that knowledge-based systems are a key enabling technology for this integration. Finally, Blomqvist [231] also strongly advocate for semantic technologies as a means to provide integration and intelligence for decision support.

### 2.1.9 DELIVERING 'SMART' THROUGH ARTIFICIAL INTELLIGENCE AND SEMANTIC TECHNOLOGIES

In order to provide the most valuable decision support possible, software should derive insights from the data which are aligned with the target business processes. This is central to the 'smart' of smart cities, and draws on many fields of research. This section briefly provides an overview of some of advanced applications from these fields, and builds an argument for the need to support these applications through IoT and semantic technologies in order to realise their potential.

A spectrum of intelligent computing is provided by Sheth et al. [232], who compare concepts such as artificial intelligence, ambient intelligence, and cognitive computing, where the difference is the degree of human-centricity. Artificial intelligence has also evolved for competency at specific tasks, such as self-driving cars. The role of artificial intelligence in Smart Cities has been broadly noted [23], [26], [30], [233]–[236], and is often accompanied with semantic technologies.

Some of the (not mutually exclusive) fields of research which may be leveraged by smart cities include optimisation, artificial intelligence, cognitive computing [236], machine learning [237], fuzzy systems [234], neural networks [235], and agent-based systems [238]. Identifying all promising avenues of research in this area is out of scope, but multi-agent systems are considered especially pertinent.

The use of agency and distributed intelligence is a significant trend within smart cities, which involves a virtual network of intelligent and autonomous controllers (modelled as software agents). This allows the management intelligence to be modularized and hence more adaptable, resilient and scalable than in centralized approaches [239]. In a multi-agent system approach, complete knowledge of the system is not required at any individual node, but each system component acts autonomously towards a set of predefined goals to optimize the overall system's performance [240]. Software agents can interact and communicate with their environment and with other agents via predefined interfaces. The behaviours of agents are conditioned by their individual goals, which can be in cooperation or in competition with the goals of other agents. The behaviour of the overall system then emerges as a result of the behaviours of its agents. By designing the agents, their interactions and their goals carefully, this emergent nature can be leveraged to optimize the performance of the overall system. In order for this emergent behaviour to manifest properly, it is imperative that agents are able to communicate effectively, which has led to the widespread use of ontologies in agent-oriented programming, as a common vocabulary for agents to use in exchanging messages.

As each agent autonomously acts with the knowledge available to it, the failure or introduction of components or communication pathways does not cause total system failure, leading to the approach's powerful resilience through adaptability [31], [88], [91], [241]–[245]. Further, as intelligence and computing power is provided at each agent, the approach is more scalable than centralized control as the computing power available will increase alongside the complexity of the system. This paradigm suits the nature of IoT well, as a core feature of IoT is its highly distributed and heterogeneous nature.

Given the recent acceleration of IoT, this represents a transformative data ecosystem which could be leveraged by artificial intelligence, if issues around interoperability, scalability and veracity are overcome. The need for convergence between the fields of information systems and artificial intelligence was noted before 1990 by Brodie [246], but he stated in 2011 that almost no progress had been made [247], and this inertia continues to the present day.

### 2.1.10 KNOWLEDGE EXCHANGE AS A CRITICAL ENABLER OF LEVERAGING SYSTEM THEORIES

In order to deliver advanced applications which integrate heterogeneous systems, it is required to align the syntax, data formats, semantics, logic, and overall domain perspectives, of the constituent systems' data. This mandates a comprehensive, powerful, and seamless knowledge management solution in order to achieve truly smart cities. Such knowledge management clearly requires highly effective methods of creating, exchanging, and utilising system knowledge. As an extension of systems theories within the modern era of ICT, there has been much progress towards capturing systemic knowledge in formal, machine-interpretable descriptions of cities [46], [231], [248]–[252]. This is a critical requirement for the leveraging of system theories, as the application of intelligent system management approaches mandates communication between agents which is seamless, powerful, and comprehensive [208], [253].

This need for powerful communication is a core benefit of the semantic web, and can be observed in literature regarding communication theory [43]. Specifically, the role of communication can be divided into three parts; the composition of the message from an agent's conceptualisation of an observation, the exchange of that message between agents, and finally the 'consumption of' and 'value derivation from' the message by another agent. Given this understanding, the role of interoperability, and semantics, in facilitating this process is clear. System components must be able to do more than simply exchange data, in order to interact in the more complex ways which are required to leverage complex system theory knowledge in smart cities. Arguably, The Internet of Things does not currently consider this powerful interoperability, despite striving to address interoperability challenges. This is discussed more in Section 2.1.10, and Section 2.3 identifies that semantic technologies address this need by utilising description logic and axiomisation to capture expressive machine-interpretable descriptions of systems.

## 2.2 INTERNET OF THINGS: THE CANVAS FOR SMART

Recent years have seen unprecedented growth in wireless and mobile devices, low-cost sensors, ubiquitous connectivity, and network communication speeds. This has led to a boom in the number of web-connected devices, with an anticipated global market impact of $11.1 trillion per annum by 2025 [20]. Alongside a growing acceptance of pervasive technology, this has caused a vast change in the integration of the cyber and physical domains [254]. This is widely acknowledged around the world through the concept of an Internet of Things. Whilst such connectivity is not a new concept, the scale and nature of the interconnectivity has accelerated recently due to the increasing penetration of technology in all areas of life, fuelled by reducing costs associated with hardware, improving communication networks, and strong economic, social, and political drivers for change [255]. This has led to the coagulation of various paradigms such as machine-to-machine communication, smart everything, and industry 4.0, through the concept of an Internet of Things. The growth of this field is shown in Figure 3 and Figure 4, which indicate that academic interest in the field has grown exponentially from circa 2009, with popular interest following from circa 2013.

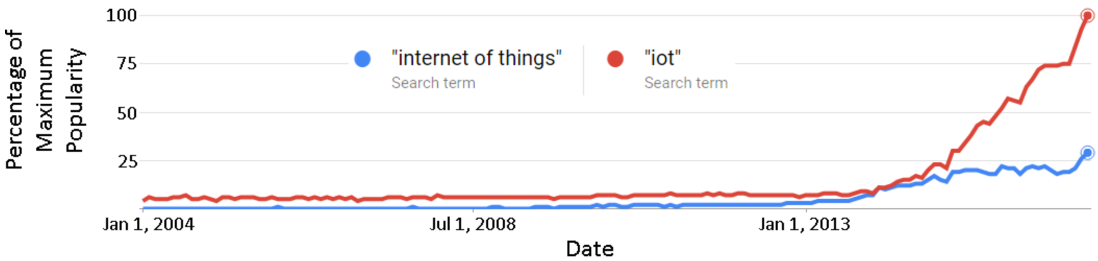


**Figure 4: Growth in Google search popularity of IoT**

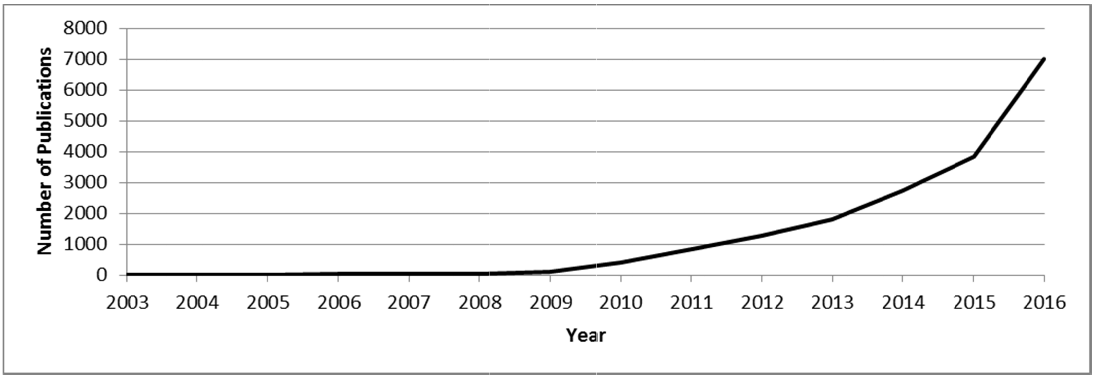


**Figure 5: Growth in academic IoT sources, based on Scopus search "internet of things"**

The most-cited academic source on the Internet of Things[1] [256] presents a clear statement of the maturity of the field in 2010, and states that the paradigm has emerged from the convergence of three visions. These complementary visions are thing-oriented, internet-oriented, and semantic-oriented, although the authors also stress the importance of social aspects. Moirandi et al. also give a seminal review of the field [254], emphasising its cyber-physical nature, and stating the key features which IoT technologies must support: heterogeneity, scale, ubiquitous wireless connectivity, energy constraints, geographic localisation of devices, distributed intelligence, semantic interoperability, and embedded security and privacy.

The role of IoT in smart cities has been well-acknowledged [18], [19], [21], [22], [257], as smart cities offer an ideal use case for many of the features of IoT, such as its wireless connectivity and heterogeneity. It is critical to note that an independent 'smart city industry' has not emerged, rather the 'smart market' aims to bring value to existing sectors, where 'smart city' aims to foster the growth of, and integrate where possible, the subdomains such as smart grid, smart water, eGovernance, smart health etc. This means that smart technology is not currently replacing existing sectorial models; rather it is seeking adoption within existing industries. However, it has been noted that industry adoption of IoT technologies has been slower than anticipated, typically due to unease at the embryonic, rapidly changing, and highly contested nature of IoT technologies as well as concerns relating to standards, security, longevity of the IoT movement, and inertia due to legacy systems and business processes. In order to deliver value from IoT technologies, organisations must be confident in the business case of adopting the new

[1] 2492 citations at time of writing, as per Scopus search results for "internet of things"

technologies, which in the short term requires high-level endorsement, successful examples of innovation, and accessible step-change solutions.

This section constructs an argument for the existence and significance of the gap in IoT which this thesis aims to address. Firstly, the section offers an overview of IoT platforms and interventions in smart cities, followed by clearly identifying a key challenge in IoT technologies, before describing the existing work in this space, and finally some further considerations which must be considered, and are relevant to the application layer challenge observed.

### 2.2.1 LEVERAGING IOT IN SMART CITY INTERVENTIONS

Some of the key domains, or verticals, where IoT provides value in smart systems are energy, water, environmental monitoring, mobility, health, governance, waste, food, manufacturing, agriculture, wearables, homes, and other buildings, to name a few [18], [22], [257], [258]. IoT will also enable transformative change in information systems and business intelligence applications, with impact across all domains. The potential value of IoT has been broadly disseminated, so this section will only briefly touch on this before progressing to the primary identified challenge.

The UK government has identified 5 target sectors for IoT value: transport, energy, healthcare, agriculture, and buildings [22]. AIOTI has identified 10 IoT domains, which includes those targeted by the UK government [258] and points to some example research. Within the smart health domain, AIOTI points to the BUTLER SmartHealth trial, and the iCore pilot at Trento hospital, which aims to track portable equipment locations and usage for predictive capabilities. RFID technology has been especially touted for such applications, and is a key enabling technology for IoT [256]. Other medical IoT use cases include patient-sensing (both in-patient and out-patient) [256], community-based pervasive healthcare [259], and ambient assisted living [260].

A great deal of literature has considered the application of IoT in smart grids [261]–[264]. Karnouskos emphasised the cyber-physical nature of this paradigm [261], and also highlights the business value of this evolution of energy systems. Yun and Yuxin describe their perception of IoT features in the context of smart grids: comprehensive sensing, reliable transmission, and intelligent processing [263]. They provide a brief overview of some applications of IoT in smart grids, primarily in terms of telemetry and systems integration. Bui et al. provide a more detailed analysis of smart grid use cases [262], such as asset management, network performance monitoring and reporting, system reconfiguration, and integration of DERs. Whilst these works offer some basic context, IoT is likely to transform the energy industry, in part through demand-side management [83], as IoT could be a key enabler for integrating smart meters, in-home displays, consumption sensors and domestic DERs.

### 2.2.2 IOT PLATFORMS IN SMART CITIES

Two groups of IoT research in smart cities can be identified: platforms which support IoT interventions, and action research alongside industrial partners to implement and observe IoT in their systems. This section described work on the former, with the latter group discussed in the following section.

Many examples exist of IoT platforms aiming to coordinate data management in smart cities [265]–[268]. The CityPulse project [269] emphasises scalable IoT stream processing, and includes semantic tagging of streams, but this is based only on a simple ontology describing the domain of data and event streams, rather than contextualising the data through a model of the target socio-technical system. The ALMANAC project [266] proposed a service oriented architecture for the collection and analysis of near real time information, and again boasted semantic interoperability. The ALMANAC platform went beyond the semantic modelling conducted in CityPulse to include domain concepts, such as their ontology for water applications [270], but this only described 6 types of object, so again lacked domain contextualisation. The RERUM project again proposed an IoT framework, but emphasised its security and privacy aspects [267]; this again utilised a semantic model, but only described the cyber-physical nature of implemented IoT systems, and not the underlying socio-technical system. As well as IoT platforms, the role of standards in enabling interoperability is critical, and one effort which accomplishes this in an accessible and open manner, whilst incorporating semantic extensibility, is the Hypercat standard [47].

**Table 1: Summary of previous IoT platforms and projects**

| Project Acronym | Description | Keywords |
|---|---|---|
| **CityPulse [269]** | Combines a knowledge-based approach with reliability testing to provide a platform-based smart city IoT applications. | IoT, platform, Analytics, Big Data |
| **ALMANAC [18]** | IoT platform which collects, aggregates, and analyses real-time or near real-time data from heterogeneous sensors and actuators to support Smart City processes. | IoT, platform, semantic interoperability, SOA |
| **RERUM [267]** | Smart city IoT platform with an emphasis on security, and coping with heterogeneity. | IoT, resilience, cybersecurity |
| **VITAL [271]** | Heterogeneous IoT system and service integration project. | IoT, platform, applications, interoperability |
| **TRESCIMO [272]** | Test beds for IoT innovation in Europe and in South Africa. | IoT, M2M, resilience |
| **ClouT [273]** | Cloud Computing project aiming to bridge IoT with 'Internet of People', emphasis on effective integration. | IoT, platform, service-oriented |
| **SMARTIE [274]** | Security, privacy and trust project for consumer IoT data. | IoT, cybersecurity |
| **FIESTA [275]** | IoT platform for heterogeneous IoT technology integration. Emphasis on a semantics-based solution and providing more than just software outputs. | IoT, interoperability, |

### 2.2.3 APPLICATION LAYER INTEROPERABILITY CHALLENGE

The Industrial Internet Consortium has recently published their view of the industrial Internet; decomposing the space into physical systems and functional domains, the latter of which is then represented as a business domain above three parallel domains: operations, information, and application, with a control domain then interfacing between these and the physical systems. The de facto internet 'hourglass' offers a technological viewpoint of the spectrum from physical to cyber domains, with technologies such as fibre and radio feeding upwards through Ethernet and Wi-Fi to reach a common technology of IP, before then spreading outwards again to HTTP, CoAP etc., and ultimately the application layer at the top. Further, the IoT World Forum has described a layered view of IoT based on information flows, from physical devices through to a business layer via the layers of connectivity, edge computing, storage, data abstraction, and applications [276]. These models are loosely aligned with the perspective adopted in this report and presented in Figure 5 below.

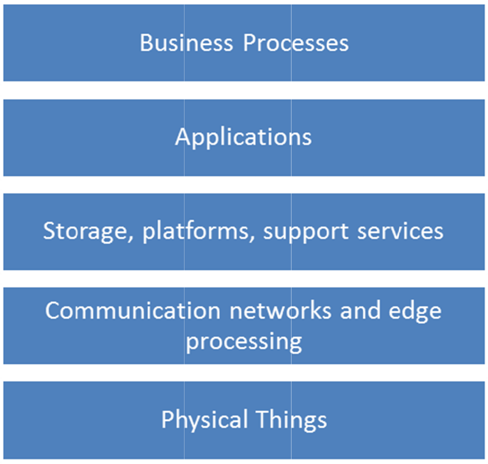


**Figure 6: Reference model of IoT systems, aka the 'IoT stack'**

Of the layers shown in Figure 5, this thesis focuses on interoperability amongst applications and IoT platforms. This is a key enabler of IoT innovation, as this location in the 'IoT stack' is often an interface between businesses, such as where a utility acts as a data provider by web-enabling its existing systems, and an IT company then builds services and applications across these data. Further, interoperability has been touted to account for 40-70% of the value of IoT [20].

As an adaption of the W3C semantic web stack and ongoing web of things work [277], some key requirements for interoperability at this layer are:

- Resource discovery
- Resource interaction interface
- Security and privacy
- Permission and constraints on resource usage (licensing)
- Consumable response format (syntax)

- Context, meaning and provenance (semantics) of the resources and data
- Trust of all of the former aspects

### 2.2.4 OVERVIEW OF THE EMERGING STANDARDS LANDSCAPE

IoT interoperability standards are being defined by the industrial alliances and standards development organisations outlined in Table 2. The relevant application layer offerings from these organisations are summarised in Table 3. Table 3 lists the most pertinent offerings, rather than IoT standards such as IP, MQTT, CoAP etc., as these are only relevant to the lower IoT layers, so are complementary to application-layer standards. It is important to observe that 9 of 14 of the offerings have been updated since June 2016, and IEEE states that 80 of its standards are related to IoT, which highlights the pace and relevance of the IoT interoperability space at present.

**Table 2: Key IoT alliances and standards development organisations**

| **Organisation Name** | SDO[2] | Total Members[3] | Key Members | Comments |
|---|---|---|---|---|
| **AIOTI** | | | Cisco, IBM, Intel, Samsung, Vodafone, Philips | European focus |
| **AllSeen Alliance** | | 137 | Honeywell, Microsoft, Panasonic, Sony, Qualcomm | Open source code-base; de-facto SDO |
| **Hypercat Alliance** | | 1100[4] | Flexeye, IBM, Intel, Cisco, BSI, KPMG | Application layer and discoverability focus |
| **IIC** | | 245 | GE, IBM, Cisco, AT&T, Intel | Guiding IIoT SDO efforts |
| **IPSO Alliance** | | 38 | Google, Oracle, Bosch, Intel, Texas Instruments | Protocol and data layer focus |
| **Thread Group** | | 215 | Google, Dell, LG, Samsung, Qualcomm | Domestic use cases only, networking focus |
| **IETF** | ● | | | |
| **ITU-T** | ● | 268 | | Has specific IoT group (SG20), part of ITU, focused on smart cities |
| **OASIS** | ● | 278 | IBM, Cryptosoft, Microsoft | Information modelling focus |

2 Standards development organisation

3 Based on publicly displayed information at time of writing

4 Individual members

| | | | | |
|---|---|---|---|---|
| **OCF** | ● | 224 | Intel, Qualcomm, Cisco, Microsoft, GE Digital, IBM | Sponsors IoTivity and includes UPnP, application layer focus |
| **OGC** | ● | 520 | IBM, Oracle, Google, Airbus, Bentley | Geospatial data focus |
| **OMG** | ● | 268 | Microsoft, Airbus, AT&T, IBM, NASA, W3C | Information modelling focus |
| **One M2M** | ● | 237 | ETSI, CEN, Intel, Cisco, IBM | |
| **OPC** | ● | 442 | Microsoft, IBM, Bosch | Industry 4.0 focus |
| **W3C** | ● | 417 | Google, Intel, IBM, Cisco, Oracle, Microsoft, Apple, SAP | |

**Table 3: Key application layer interoperability and discoverability offerings**

| Offering Name | Organisation | Endorsement | Last known update |
|---|---|---|---|
| **Hypercat Standard** | Hypercat, BSI | Approved | Jun-16 |
| **WoT, TD** | W3C | Draft | Sep-16 |
| **Semantic web standards** | W3C | Approved | Mar-13 |
| **SensorThings API part 1** | OGC | Approved | Jul-16 |
| **Core Framework** | OCF | Draft | Dec-15 |
| **XMPP-XEP-0347** | XSF | Experimental | Nov-15 |
| **AllJoyn Framework** | AllSeen Alliance | N/A | Aug-16 |
| **CoRE resource Directory** | IETF | Draft | Jul-16 |
| **Service Layer Core Protocol Specification** | OneM2M | Approved[5] | Aug-16 |
| **Web-Enabled DDS** | OMG | Adopted Beta | Apr-15 |
| **CAMP** | OASIS | Draft | Sep-16 |
| **Weave, Thread, Brillo** | Google | N/A | Oct-16 |
| **Bip** | WOT.io | N/A | Dec-15 |
| **Zetta** | Apigee | N/A | Oct-16 |

The AllSeen Alliance and the Open Connectivity Foundation (OCF[6]) are especially pertinent organisations, although these mainly address interoperability at the lower layers of the 'IoT stack'. This contrasts the Hypercat approach, which enables interoperability at the application level. Historically, the OCF work directly rivalled that of the AllSeen Alliance, although their recent union changed the landscape significantly. OneM2M also aims to offer a unified end-to-end IoT solution, and has published approaches for integration with both the OCF and AllSeen standards.

[5] Not yet approved by 'Type 1' partners

[6] The OCF is the recently rebranded Open Interconnect Consortium, and now includes the ubiquitous UPnP standard, and the IoTivity open source reference implementation of its specifications.

However, along with OASIS, these organisations all primarily focus on lower layer communication.

Google is not directly supporting these initiatives; instead it is focusing on its Weave, Brillo, and Thread offerings, primarily for smart home. Whilst the Thread group has recently allied with OCF, the nature of this relationship remains to be seen. It is relevant though that Google is a member of W3C, which has recently allied with the Hypercat consortium to progress a shared vision. The semantic web standards of W3C such as SPARQL and RDF are not considered as IoT standards, but are highly relevant to this thesis. Another major market player, Apple, is not contributing to these open efforts and is focusing on inward innovation in its HomeKit products, directly rivalling Google. Also, the CAMP standard of OASIS is limited to tasks such as deploying, stopping, starting, and updating applications, rather than resource discovery, so is not considered further.

Four of the offerings in Table 3 are open source code bases, including the AllJoyn Framework, and the significantly smaller Bip and Zetta offerings, which are still relevant albeit not likely to represent competition for AllJoyn. As open source projects, none of these would undergo a formal standardisation process, which weakens their credibility somewhat.

### 2.2.5 RELEVANT LOWER LAYER INTEROPERABILITY STANDARDS

This section briefly introduces the aspects of key comprehensive IoT standards relevant to application layer interoperability. This omits many IoT standards which exist only at the lower layers of the IoT stack, such as MQTT, HTTP etc., which are not directly relevant to this thesis. Firstly the relevant aspects of the end-to-end solution by OneM2M are addressed, followed by the relevant aspects of the work by OCF and the AllSeen Alliance, which all emphasise the lower layer machine-centric aspects of interoperability.

OneM2M aims to provide an end-to-end solution for IoT interoperability; it only provides guidance on using the end-to-end solution to address this problem. Application layer interoperability is mentioned in the OneM2M application developer's guide TR-0025 [278], but this functionality is described in the context of end-to-end OneM2M systems, an unlikely situation in most scenarios.

The OneM2M TS-0004 specification [279] defines communication protocols, data formats, and interfaces for applications, based on the architecture specified in TS-0001 [280], although this operates at the middle of the OneM2M 3-layered stack of ‘network’, ‘common services’, ‘applications’. Further, the resources considered are primarily machine resources such as control policies, schedules, and memory, rather than application domain ‘Things’ such as sensors, cars, buildings. The current suite of OneM2M standards are situated alongside the AllJoyn and OIC offerings; closer to the communication and networking levels of interoperability, rather than the application layer.

The OCF has drafted a core specification that aims to achieve “resource-based interactions among IoT artefacts”, based on a RESTful API, such that each physical object has a URI and an API for interaction. The standard also defines endpoint and resource discovery, primarily based on the CoAP protocol. OIC resource type definitions include the properties Name (human readable; mandatory), location (JSON object which contain lat, long; optional), location name (human readable; optional), currency (optional), and region (optional). OCF defines 3 types of resource discovery: direct, indirect, and presence based, where only direct discovery is mandatory. Given that remote discovery is not mandated by the OCF standard, it is not well suited for the application layer challenge identified. Also, the approach is not well suited for semantic web integration due to the recommendation that RAML and JSON Schema be used for defining resource types.

The AllSeen Alliance has produced the AllJoyn Framework; a body of code which facilitates the development of applications which can discover and interoperate with varied nearby devices. Whilst perceived in a similar space to application layer interoperability, this framework differs fundamentally in that it is intended for local networks, where data is transported directly by Wi-Fi, Ethernet etc. The solution does offer a ‘Gateway Agent’ which aims to support cloud-based use cases, but this is supplementary to the core objective. Given this, this offering does not aim to address the challenge identified. The AllJoyn Framework supports Thing discovery through self-announcement on a local network, and includes an ‘interface definition’ which specifies the properties which should be shared, such as ‘friendly names’, ‘supported languages’, and ‘service port number’. Knowledge of these properties may be useful in further defining the W3C Thing Description model.

### 2.2.6 APPLICATION LAYER DISCOVERABILITY AND INTEROPERABILITY

This section describes the standards and initiatives surrounding application layer interoperability, and discusses the relevance and merits of each offering. Firstly, the BSI Hypercat standard is described, followed by the W3C Web of Things framework. Offerings from OGC, XMPP, and IETF are then discussed.

#### 2.2.6.1 HYPERCAT

Hypercat aims to facilitate application layer interoperability to accelerate collaborative IoT innovation. The Hypercat standard [47] provides RDF-like mechanisms for this in a highly accessible, lightweight, flexible and extensible manner, and has been adopted by BSI following its development through 2 Innovate UK projects and collaboration through the Hypercat Alliance. The Hypercat standard describes 3 aspects: i) a mandatory data format, ii) an API specification, and iii) other suggested extensions for common use cases.

The central tenet of Hypercat is the mandatory file format, which requires that a Hypercat server must return a catalogue of resources in a specific JSON format. The Hypercat standard also suggests a REST style API, and an example key-based authentication method. The standard also suggests extensions to this such as subscription, further security options, various search methods, a means to integrate Hypercat into the linked data and semantic web ecosystem further, and finally a method for describing the usage licence under which a described item is provided.

Hypercat is being leveraged across UK cities such as London, Bristol, Milton Keynes, and Manchester. The recent announcement of the use of Hypercat in the £10M smarter Manchester project ‘CityVerve’ shows that Hypercat has ongoing support. This, coupled with the adoption of Hypercat by BSI, and the recent agreement to collaborate with W3C towards a shared vision, leads to the recommendation that Hypercat represents the best modular approach to application level IoT interoperability at present.

#### 2.2.6.2 WEB OF THINGS AND THING DESCRIPTION MODEL (WORLD WIDE WEB CONSORTIUM)

The W3C Web of Things Interest Group was launched following a workshop in June 2014 which was hosted in the context of the COMPOSE EC FP7 research project, and has recently become a full W3C working group. The group has published a draft WoT architecture, and a draft document of WoT "current practices" [277], although these remain very much 'works in progress'. These include a "Thing Description" model, and a "Scripting API". The W3C "Thing Description" model is an RDF based description of 'things', and uses a lightweight vocabulary to describe a thing's metadata, security, communication methods, properties, actions, and events. The WoT work is based on the existing standards of W3C, as is the Hypercat solution, so they are compatible. These models are compared in Figure 6.

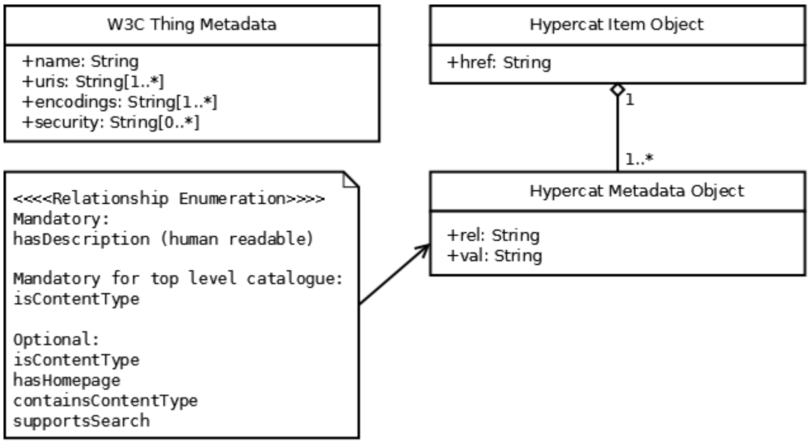


**Figure 7: Comparison between WoT Thing Metadata and Hypercat Item descriptions**

#### 2.2.6.3 SENSORTHINGS API (OPEN GEOSPATIAL CONSORTIUM)

The OGC SensorThings API [281] is an IoT-specific development of their Sensor Web Enablement (SWE) standards, and is split into two parts: sensors and actuators, the latter of which has not yet been published. This frames the IoT in a manner which ignores virtual 'things', such as BIM models, virtual resources, and 'soft' sensors. The alignment with the OGC observations and measurement (O&M) standards may be beneficial to environmental reporting agencies which are already familiar with the model, but may be seen as unnecessary complexity to IoT developers with different use cases. The data model of the SensorThings API is shown in Figure 7.

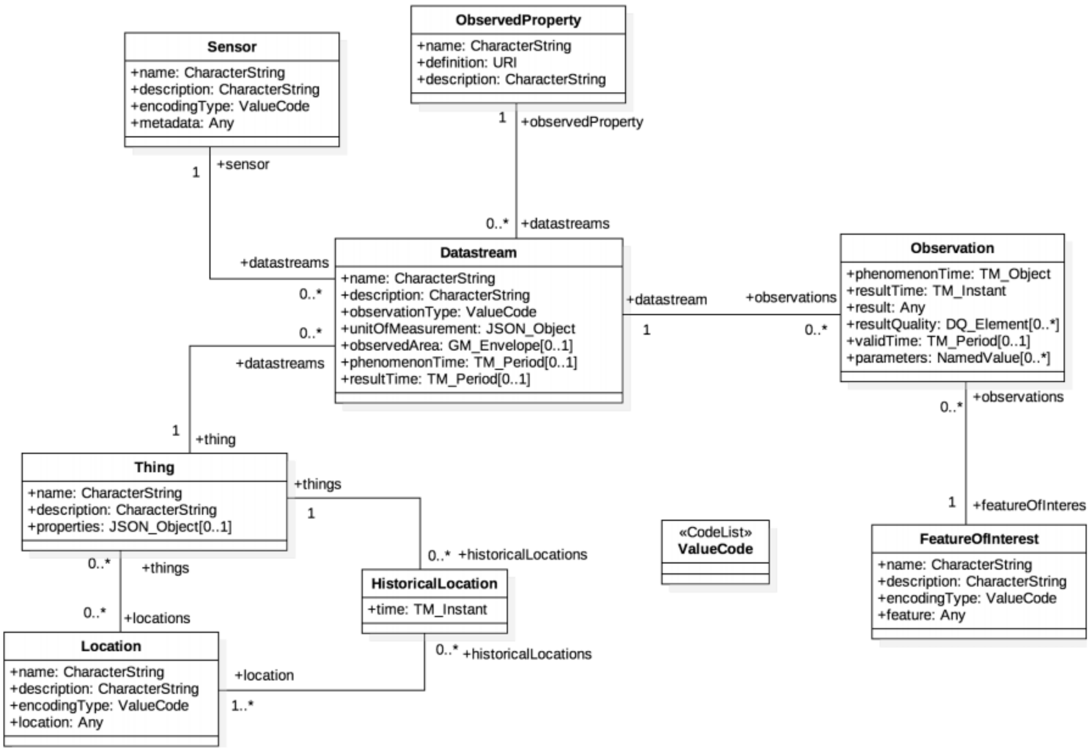


**Figure 8: SensorThings API entities**

Only a subset of the properties are mandatory in describing a thing: local id, selfLink (URL), navigationLink (URL), name, description (human readable), and properties. The approach assumes that sensor data will be retrieved via the same service which provides discoverability, which may be unsuitable. The approach doesn't lend itself to semantic web integration, due to the less expressive entity-relationship modelling.

#### 2.2.6.4 XMPP-XEP-0347 (XMPP STANDARDS FOUNDATION)

The XEP standards are extensions of the Extensible Messaging and Presence Protocol (XMPP), standardised by the XMPP standards foundation. XMPP-XEP-03467 is an extension to support IoT discovery. The standard is more focused on the installation and configuration phase of IoT Things, and includes 17 use cases surrounding the topic of IoT discovery, including the finding of a 'thing registry', and searching for public things in a registry. The search functionality accepts XML serialised tag-value search terms. The reliance on XML serialisation is detrimental to usage in IoT due to verbosity, with Hypercat offering a more lightweight approach. The reliance on tag 'codes' such as 'LAT', 'LON' is also poorly aligned with technologies, as the use of a URI predicate would leverage well accepted W3C standards better. The snippets below show an example search query and response from a Thing Registry respectively [282]:

Search query:

```
 <iq type='get'
       from='curious@example.org/client'
       to='discovery.example.org'
       id='9'>
      <search xmlns='urn:xmpp:iot:discovery' offset='0' maxCount='20'>
          <strEq name='MAN' value='www.ktc.se'/>
          <strEq name='MODEL' value='IMC'/>
          <strMask name='SN' value='39487*' wildcard='*'/>
          <numRange name='V' min='1' minIncluded='true' max='2' maxIncluded='false'/>
          <numRange name='LON' min='-72' minIncluded='true' max='-70'
maxIncluded='true'/>
          <numRange name='LAT' min='-34' minIncluded='true' max='-33'
maxIncluded='true'/>
      </search>
   </iq>
```

Query response:

```
<iq type='result'
       from='discovery.example.org'
       to='curious@example.org/client'
       id='9'>
      <found xmlns='urn:xmpp:iot:discovery' more='false'>
          <thing owner='owner@example.org' jid='thing@example.org'>
             <str name='SN' value='394872348732948723'/>
             <str name='MAN' value='www.ktc.se'/>
             <str name='MODEL' value='IMC'/>
             <num name='V' value='1.2'/>
             <str name='CLASS' value='PLC'/>
             <num name='LON' value='-71.519722'/>
             <num name='LAT' value='-33.008055'/>
          </thing>
          ...
      </found>
   </iq>
```

#### 2.2.6.5 CORE RESOURCE DIRECTORY (INTERNET ENGINEERING TASK FORCE)

The CoRE Resource Directory of IETF is a working document which aims to support remote discovery of resources in M2M applications, by defining interfaces for CRUD functions of a Resource Directory. A Resource Directory contains a set of endpoint descriptions, which are locations through which resources can be accessed, and which can be conceptually aggregated into groups and domains, as shown in Figure 8. The approach only requires that endpoint names be unique within a domain, which contrasts the URI-based approach of Hypercat and W3C. The response format following a GET request to a Resource Directory is shown in Figure 9.

```
+------------+
|   Domain   | <-- Name
+------------+
     |     |
     |   +------------+
     |   |   Group    | <-- Name, Scheme, IP, Port
     |   +------------+
     |     |
+------------+
|  Endpoint  |  <-- Name, Scheme, IP, Port
+------------+
      |
      |
+------------+
|  Resource  |  <-- Target, Parameters
+------------+
```

**Figure 9: IETF resource directory information model**

```
The following example shows a client performing a resource lookup:

Req: GET /rd-lookup/res?rt=temperature

Res: 2.05 Content
<coap://[FDFD::123]:61616/temp>;rt="temperature"

The following example shows a client performing an endpoint type
lookup:

Req: GET /rd-lookup/ep?et=power-node

Res: 2.05 Content
<coap://[FDFD::123]:61616>;ep="node5",
<coap://[FDFD::123]:61616>;ep="node7"
```

**Figure 10: IETF response format to resource and endpoint lookups**

Overall the IETF approach offers some advantages in constrained IoT environments, but isn't well suited to application layer and platform layer interoperability, where human readability would be favourable, and a standardised encoding, like Hypercat's JSON, would be a benefit.

## 2.2.7 DATA QUALITY, SECURITY AND PRIVACY IN IOT

A critical barrier to the adoption of IoT is the perceived security and privacy risks [22], and IoT is at the heart of many of the security and privacy concerns around smart cities. This section briefly reviews some of the specific concerns, measures, and gaps in mitigating this, as well as briefly considering data quality.

A great deal of relevant literature has emerged recently [283]–[291], for example the good discussion regarding IoT security and privacy concerns by Atamli and Martin [287]. They identify the following threat sources: malicious users, bad manufacturers, and external adversaries. They identify the following attack vectors: device tampering, information disclosure, privacy breach, denial-of-service, spoofing, elevation of privilege, signal injection, and side-channel attacks. Gayathri et al. [283] outline their perceived attack vectors: physical attacks, side channel attacks, cryptanalysis, software attacks, network attacks, disclosure of data, breaching privacy, denial-of-service, spoofing, privilege elevation, and signal injection. Nurse et al. consider a different perspective of attack vectors, based on insider threats [285], where they provide a detailed account of 16 modified attack vectors. Ren et al. adopt a very interesting approach to attack vector descriptions, using ontological modelling [292].

Fremantle et al. observe that traditional role-based security is poorly suited to IoT, due to the number of devices, the distributed nature of control and agency, and the need for privilege delegation [286]. Another critical consideration is that integrating IoT into existing systems results in more complex and frequently changing systems, adding security issues [283]. These challenges are, broadly, still open issues, so the present work will endeavour to use best practice.

Cybersecurity fundamentally aims to control access to information, services, and physical facilities, through concepts such as authorization and authentication [293]. (ISC)$^2$, the leading cybersecurity consortium, has stated 10 best practices for secure software development [294]:

1. Protect the Brand Your Customers Trust

2. Know Your Business and Support it with Secure Solutions

3. Understand the Technology of the Software

4. Ensure Compliance to Governance, Regulations, and Privacy

5. Know the Basic Tenets of Software Security

6. Ensure the Protection of Sensitive Information

7. Design Software with Secure Features

8. Develop Software with Secure Features

9. Deploy Software with Secure Features

10. Educate Yourself and Others on How to Build Secure Software

These overarching practices should be specialised for the unique challenges which IoT presents [295]. Fremantle et al. restate this, and propose the use of OAuth for scalability alongside MQTT and TLS [286]. They also recommend standardization of OAuth2 usage mechanisms for IoT, but identify the scale of work required for this. Nurse et al. recommend future research towards i) better documentation of attack vectors, ii) investigation of how mobile device practices can be extended for IoT, iii) investigation of business societal policies, and iv) ethical and legal issues [285]. Vasilomanolakis et al. [291] review security measures taken in several leading edge IoT architecture research projects, which are summarised in Table 4 below. This demonstrates the emphasis to date on lower-layer security, rather than device and application security, and resilience and privacy issues. Abomhara and Køien note the recent advances in access control, around the "so-called usage control", which allows more fine-grained control over thing usage, and a spatial approach which captures the geographical nature of threats in highly distributed systems [288].

**Table 4: Security measures taken in several leading edge IoT architectures [291]**

| Project | Security Measures |
|---|---|
| IoT-A | **Network security**: Key exchange & management component, using IP Security tunnels between gateways.<br>**Identity management**: Authentication, authorization, and attribute-based access control modules.<br>**Also**: Pseudonymisation module, trust and reputation module (thing trust only, not data), fault handling model |
| BeTaaS | Separate mechanisms for each layer of its IoT stack.<br>**Network security**: Key management component, with certificate authority. Also; directory service for cross-organisation scenarios, and elliptic Curve Cryptography for constrained devices<br>**Identity management:** Authentication component, with gateway and application scenarios, based on OAuth. Separate authorization component. |

| | |
|---|---|
| | **Trust:** Dedicated component which aggregates metrics about security mechanisms, QoS, and battery load etc.<br>**Also:** Resilience component |
| OpenIoT | **Network security**: HTTP with TLS protocol, also allowing for IPSec tunnels<br>**Identity management:** Central service based on OAuth, with authorization based on the RBAC model.<br>**Trust:** Trust module correlates proximal sensors to produce trust labels, although full method is unclear.<br>**Also:** Resilience is approached through an inventory of things, which restructures the connections in the event of failure. |
| IoT@Work | **Network security:** Extensible Authentication Protocol for lower-layer security<br>**Identity management:** Authentication is provided by network security, authorization is provided by Capability-Based Access Control.<br>**Resilience:** Network slice approach, which virtualises network links and promotes robustness. |

It is important to recognise that a one-size-fits-all solution to security is not appropriate; every organisation and use case warrants a unique consideration. It is also important to recognise that IoT-enabled does not imply openness to any extent, it only states a paradigm adoption and implies a set of technologies to choose from in system design.

### 2.2.8 INTEGRATING LEGACY SYSTEMS IN IOT SYSTEMS

Whilst IoT may hold significant value, the transition period of adopting these technologies must be well considered, which requires a robust means of integrating existing and outdated systems with modern IoT technologies within businesses [296], [297]. This section briefly discusses some of the challenges and latest solutions in this space.

At a high level, Khoshafian outlines 5 key challenges in modernising legacy systems [298]:

- IT systems grow in an incremental manner over time
- They represent very significant investments
- They are typically difficult to change
- The majority of IT budget is used on mandatory routine operations
- Upgrading legacy systems typically consumes more than 75% of IT spend

Beyond this list, Gaiser et al. state that the challenges include maintaining security, and ensuring scalability [299] and they propose a set of best practices, including:

- Start with a dedicated team
- Prioritise requirements engineering
- Understand the sensor data to be collected
- Design the network infrastructure
- Check the operating environment
- Enshrine cybersecurity principles
- Plan for scalability
- Ensure a support model is in place

These sources provide useful background context to this important challenge, and acknowledges that integrating legacy systems with the latest technology has been an ongoing issue for decades. Similar challenges and rhetoric to that surrounding IoT integration are observed in the literature regarding earlier technology innovations, such as when service-oriented architectures were the 'latest technology' [300], and even earlier at the popularisation of Java [301].

As IT and OT technologies are widely varied across industries and all permutations cannot be covered here, the case of integrating IoT with SCADA systems will be considered further, as a significant issue in utilities and mass transit organisations, amongst many others. Sajid et al. offer a brief review of IoT-SCADA system security and identify current challenges [13]. Fahrion makes a number of recommendations towards an industrial IoT beyond SCADA: separate data producers and consumers, nurture a collaborative ecosystem, use edge processing to upgrade data to information, adopt IoT protocols, and adopt overlay networks [297]. He also advocates for the use of semantic data models, as they empower application developers. This role of semantic models is also noted by IEEE within smart grids [15]. These concepts of overlay networks which provide rapid time to market without disrupting existing processes and using semantic models to integrate aspects of IT and OT with advanced applications, is embodied throughout the present thesis.

## 2.3 DELIVERING INTELLIGENCE: THE ROLE OF SEMANTIC TECHNOLOGIES

The internet has been a revolutionary technology in all aspects of modern life, since its first conception some 30 years ago by Sir Tim Berners-Lee, based on the use of Uniform Resource Identifiers, HTML, and HTTP [302]. The introduction of user-generated content such as through Wikipedia or Facebook, caused a shift in the paradigm of the internet from being information serving, to a platform for connectivity and exchange of ideas, encapsulated through the term ‘Web 2.0’ [303]. This has since been superseded in research through web 3.0 and semantic technologies [304], which was again proposed by Sir Tim Berners-Lee and W3C, and aims to better define the meaning of information on the web.

The semantic web adds abstraction layers above existing web technologies to enable machines to understand the meaning and context of content. This is achieved in practice through the Resource Description Framework, Web Ontology Language, SPARQL Protocol and RDF Query Language, and Semantic Web Rule Language. Adding this layer of abstraction above data enables many more advanced applications, and simplifies application development by decoupling the data collection and representation from the building of software which derives value from it. However, despite the capabilities of the semantic web, it has not enjoyed the success of previous web iterations [305], which Ismail and Shaikh have discussed very recently [306]. They propose that the semantic web needs to embody the same core traits of the original web: heterogeneity, distributed and crowd-sourced, and user-centricity.

The semantic web community has fostered the development of standards and technologies which allow better cooperation between human and machine intelligence, which is at the core of the smart city value proposition. Further, it has been broadly stated that more research is needed on application layer interoperability in the IoT field [23], [307], which the semantic web is well placed to address. Therefore, the role of this important technology is considered in this section, including a consideration of the value it may bring, and challenges and opportunities to its integration with IoT, smart city, and advanced application concepts. Firstly, an introduction is given to semantic modelling, before reviewing some existing progress towards integrating this with IoT concepts. Next, this approach is compared to traditional alternatives, and is considered alongside the linked data and big data fields. A gap is then proposed of ontological modelling of emerging smart domains, and so ontology types and engineering practices are reviewed.

### 2.3.1 INTRODUCTION TO SEMANTIC MODELLING AND ONTOLOGIES

This section extends the introduction in section 1.3.3. An ontology is commonly defined as 'a formalised conception of a domain' [49]. 'Formal' means machine readable based on accepted syntax [50], such as OWL. 'Conception' refers to an ontology reflecting a world view, by providing structure to its concepts [308]. 'Domain' refers to any bounded region of concepts, which should be determined through requirements engineering and competency questions [50]. In OWL, concepts are expressed through classes, class hierarchies, object properties, data properties, restrictions and annotations.

Ontologies are similar to object-oriented programming models in some ways, but are focused on modelling the knowledge of a domain, rather than structuring the data for a specific program, or language. This degree of separation from their usage suits ontologies well for modelling knowledge in a very open, scalable, and extensible manner [309], [310]. Also, because OWL is built for the web, it is well suited to semantically interoperate the Internet of Things. Full discussions of this history and ontology engineering are beyond the scope of this section, and the reader is directed to one of the many venerable sources for a full discussion [52], [311]–[313].

Semantic models address the issue of interoperability by creating a shared data format and understanding for a domain, as well as promoting discovery, consistency and scalability [14]. These benefits have been acknowledged in the field of semantic web technologies through the World Wide Web Consortium (W3C) standards [311], [314]. A semantic web ontology is a collection of statements about a domain, serialised in a machine-interpretable format, such as RDF/XML, or turtle. An ontology file is hence a collection of such statements, which is interpreted by a machine to produce a network of concepts.

A deployed ontology typically forms part of a system's backend, in order to provide a data store which captures meaning, contextualises data, standardises terminology, facilitates rule application and produces new knowledge beyond that which is inputted. A deployed 'ontology' can be split into two distinct parts: the domain ontology itself (called a T-box) which is applicable across all instances of the domain, and an instantiation of this (called an A-box). The union of these two components forms a knowledge base, and alongside an inference engine, a query engine and a storage capability, this composes a knowledge management system. The inference engine utilises the statements made in order to infer new knowledge, and the query engine is the method of retrieving data and knowledge.

The benefits of ontologies within the fields of engineering and computer science have been described as the following [52]:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyse domain knowledge

Another central facet of ontologies is the ability to infer knowledge through explicit statements about a domain [50], [52], [308], [315]. This describes the ability to deduce that a statement is true based on other explicit statements. This action is typically performed by a 'reasoner'; a separate piece of software than the ontology or its development environment which considers the explicit statements made and infers new knowledge based on this. The inference of knowledge also requires another central assumption within ontological modelling; the 'open world assumption'. This separates ontological modelling from object oriented programming (OOP) in that an ontology assumes 'that which is not stated is not known' [50], whereas OOP assumes 'that which is not stated is not true'.

### 2.3.2 EXISTING SEMANTIC WEB OF THINGS PROGRESS

Semantic technologies and ontologies hold a great potential to integrate heterogeneous systems, and as discussed previously, this heterogeneity is the source of both IoT's biggest strengths and challenges. This distributed, varied nature must be embraced by IoT technology to solve the application layer interoperability challenge identified and unlock the true potential of IoT. The previous section described how ontologies formalise a set of truths held by a machine to represent a domain perspective, and how this can be used to integrate systems and support semantic inference. Applying these concepts alongside IoT technologies would represent a transformative technology beyond even the predicted value of IoT, by integrating existing systems and facilitating the use of cybernetic research in practice in a manner and an extent not previously capable. At the core of the research needed to facilitate this is the convergence of IoT and semantic web technologies. Research on this matter is embryonic; IoT platforms claim to offer semantic interoperability, but they only offer shallow annotations of data streams or sensors.

Broadly speaking, the main emphasis of SWoT research typically falls into one of the following (not mutually exclusive) groups:

- proposing an ontology for modelling sensors or things [24], [252], [316]–[322]
- proposing software or software recommendations [266], [307], [323], [324]
- discoverability and semantic search [27], [319], [325]
- vision, review, and guidance on best practices [29], [277], [320], [326]
- handling semantic data in constrained environments [25], [327]–[329]

Also, the outputs of the W3C web of things [277] work group is highly relevant within this discourse, and whilst not yet a formal recommendation of W3C, served as important background and guidance for the work conducted. These sources form the most directly relevant discourse to the present investigation. However, they do not consider the integration of IoT semantics with application domain semantics, nor the role of SWoT technologies amongst legacy systems, or the full breadth of possible uses for semantic technologies. This then highlights another key gap; the lack of semantic modelling of many 'smart' domains.

The recent works which recognise the need for a Semantic Web of Things share an emphasis on powerful interoperability through ontologies and open models. Jara et al. [326] present their survey and vision for SWoT technologies, highlighting interoperability at a greater level of abstraction as an evolution of IoT, through “high-level modelling of real world entities”.

Rubio et al. [213], used an ontology-based solution for subsystem and service discovery. This described services using an ontology, and reused a small number of CityGML concepts. However, it would be beneficial to also include application domain semantics in the ontology used, and to use a standard API and response format. Gyrard et al. [28] emphasise that SWoT should be an evolution of IoT for semantic interoperability across domains, through greater application domain modelling. Wang et al. focus on sensor descriptions as an extension of the SSN ontology [330] for the IoT domain [319], and earlier, Wang et al. described IoT services and other supporting aspects [24]. The work of Su et al. regards data formats for exchanging semantic information such as JSON-LD, rather than the abstractions themselves [25]. The earlier work of Pfisterer et al. [27] also proposed a SWoT vision with an emphasis on modelling real world entities, but framed the work as an evolution of semantic sensor networks. Also, [328] adopts a novel stance on SWoT, whereby web-enabled things exchange micro-ontologies, as an extension of MQTT and CoAP, hence proposing an IoT revolution through fundamental change of the lower level technologies, rather than integrating a higher-order knowledge layer above existing approaches.

One example application of SWoT principles which acknowledged the role of higher-order knowledge management, from the water domain, is [331], where a knowledge-based system is developed for the web which enables consumption knowledge to be elicited from smart metering data. Also highly significant, the WatERP project has proposed a multi-agent system based ICT platform to enable supply and demand matching in water networks, and uses a domain ontology alongside a data warehouse to manage the solution’s data [251], although the ontology is still relatively simple compared to those utilised in other domains such as energy and building information modelling.

In considering the current state of maturity of SWoT research, the models of Barnaghi [332] and Gyrard and Serrano [323] are highly relevant. Gyrard and

Serrrano indicate that IoT will evolve into the ‘web of things’, which incorporates common application protocols, but that this will evolve further into a ‘semantic web of things’, which incorporates abstractions and formal descriptions. Barnaghi doesn’t name the levels of evolution, but indicates that above thing descriptions there must be semantic models of ‘devices, resources and data’, followed by ‘domain knowledge’, and finally ‘services and applications’. Arguably, these levels of abstraction are all sub-levels of the ‘semantic web of things’ level from Gyrard and Serrano’s model. This is important as it positions the observed gap in a tangible manner; the network and application protocols of IoT/WoT are already broadly defined or *being* defined, and it is the semantic modelling at the application layer which is particularly lacking. Specifically, whilst several ‘thing ontologies’ have been proposed, as well as some ontologies for application domains and services and business processes, this field of research is significantly sparser, and no research has been observed which integrates all of the levels of these two models. Finally, it is important to note that whilst application domain ontologies exist, there are vast gaps in the modelling of ‘smart’ domains where IoT is intended to provide value.

### 2.3.3 CONCEPTUALISATIONS OF KNOWLEDGE IN IOT SYSTEMS

A common theme in SWoT literature is the evolution from data to more useful knowledge through the addition of semantic context [19], [318], [320], [333], [334], but this is represented in varying ways, both implicitly and explicitly. One of the earlier models was proposed by Evans [333], who stated that humans convert data into information, knowledge, and finally wisdom. Very similar to the pyramid proposed by Evans are the pyramids of Gyrard et al. [320], Ma et al. [318]. Gyrard et al. describe each level of insight very briefly, and Ma et al. also label the conversion between levels. The same notion is also represented by Jin et al. [19], who show a transition from data collection, through to data processing, data management, and data interpretation, and align these levels with technologies and layers of the ‘IoT stack’.

Arguably, all of these attempts to represent flows through different stages of knowledge complexity are specialisations of basic feedback loops which are common in cybernetics, but using the language of IoT and semantics. For example, the models can be compared to Kolb’s experiential learning cycle [335] from 1984, which describes a cyclical process of: concrete experience, reflective observation, abstract conceptualisation, and active experimentation. It can be stated that the models simply aim to computer-enable as much of this cycle as possible in real time operational control of systems. Specifically, concrete experience is similar to direct sensing of an environment, reflective observation is comparable to analytics and advanced application processes, abstract conceptualisation is then decision support, and active experimentation is acting on this insight. It would be remiss to not also point to the similarities to control theory and feedback systems circa 50 years ago [336], which has served as a foundation for cybernetic research and smart systems.

### 2.3.4 COMPARISON OF SYSTEM INTEGRATION AND DATABASE APPROACHES

It is imperative to thoroughly understand the alternatives to semantic technologies integrating systems. To this extent, semantic technologies can be thought of as serving two purposes: systems integration, and data storage and retrieval. Semantic intelligence is another key feature, but this sits within the field of artificial intelligence. The former role; systems integration, stems from the ability of a knowledge base to ‘point’ to other systems and resources through URLs.

Semantic technologies also facilitate interoperability by allowing systems to use a common language and domain perspective. In this role, semantic technologies are comparable to aspects of previous enterprise application integration approaches such as enterprise service busses, and can be contrasted to hub-spoke models with centralised integration capabilities. Figure 10 and Figure 11 indicate academic interest in relevant system integration concepts over time, based on Scopus searches. These figures clearly indicate that the ESB, EAI, and SOA approaches gained interest up to 2010, but that cloud computing and IoT have since become more popular. These figures also indicate IoT has been vastly more researched than these other concepts, likely due to the many disciplines which IoT intersects with.

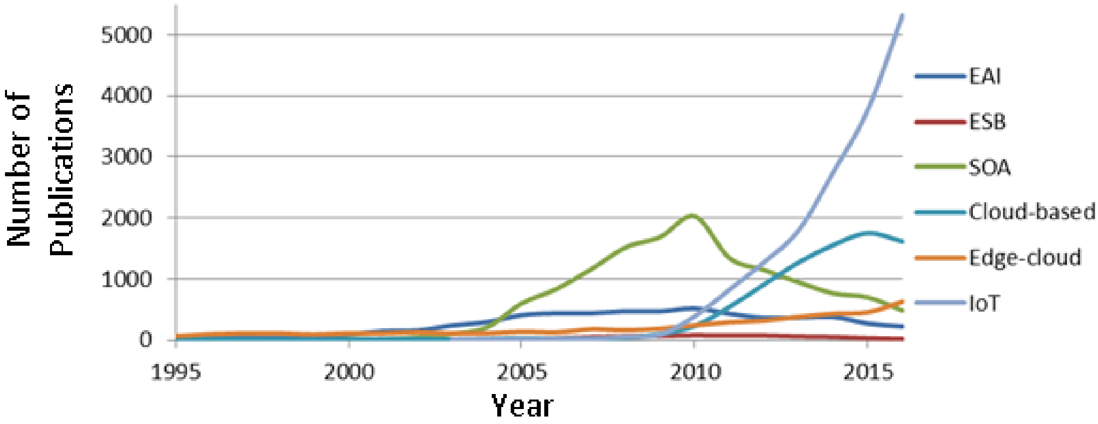


**Figure 11: Academic publications per year across systems integration approaches; absolute values**

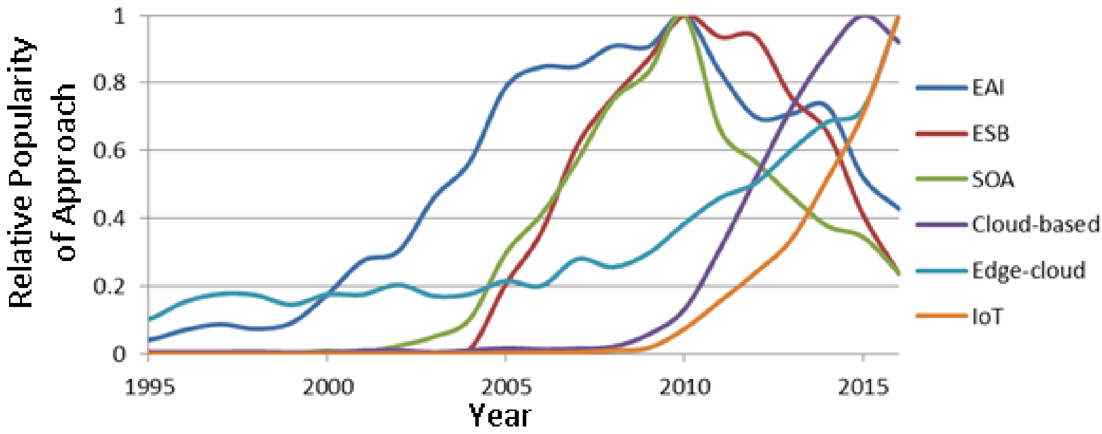


**Figure 12: Academic publications per year across systems integration approaches; normalised per approach**

Whilst a semantic web and IoT approach to systems integration is significantly different to earlier approaches, they can be complementary; such as 'semantic data warehouses' [337]. Semantic technologies have also been applied to service-oriented architectures through web service ontologies and service markup, and through a semantic services overlay above an enterprise service bus [338].

Modern semantic technologies are based on RDF date, so it is pertinent to compare RDF data stores to alternatives, such as SQL databases, NoSQL databases, data warehouses, relational databases, and graph databases. As mentioned though, semantic layers have been added to many of these [339]–[342]. The benefits and drawbacks of triple stores versus alternatives have been widely discussed, for example [343] describes how semantic technologies surpass data warehouses in modern contexts, and Ontotext defines a number of advantages of using triple stores over alternatives [344]:

- Simple to change data schemas, without downtime
- More powerful querying capabilities over distributed sources
- More standardized than other NoSQL approaches
- Easier to track provenance and data quality
- Queries are simpler than SQL
- Reasoning allows new data to be inferred from existing data
- Efficient use of resources reduce capital expenditure and total expenditure
- Simple to leverage the linked open data cloud alongside internal enterprise data

Sarnovsky compared triple stores, SQL, NoSQL, and graph databases [345], and concluded that triple stores and NoSQL databases offer competitive performances and schema flexibility, but the standard nature of SPARQL was a benefit of triple stores, whilst NoSQL support transactions and better reliability. Saikaew et al. [341] specifically compared MongoDB to Apache Jena TDB and MySQL for storing and querying RDF data, and concluded that for larger data sets MySQL performed best, due to more efficient indexing, but Apache Jena TDB performed poorly. They also concurred with Saikaew that there is a need for standardised NoSQL query languages, comparable to SPARQL. Kilintzis et al. conducted a similar comparison, between Apache Jena, MySQL, and the hybrid Virtuoso server database [346], where Virtuoso performed best, and MySQL again outperformed Apache Jena. However, the literature offers conflicting comparisons of triple stores, for example [347] calls Virtuoso the fastest and most scalable option, whereas [348] finds that Virtuoso is the slowest and least scalable option. It is a benefit of triple stores and semantic technologies then that the choice of storage and querying software is modular, so options can be tested and interchanged within a prescribed architecture.

### 2.3.5 SEMANTIC WEB AND LINKED DATA

Linked data is often referred to in the same space as the semantic web, as it is also built on the open standards of W3C, and is also based on the web and URIs. In fact the terms have been referred to as 'quasi-synonymous' [349]. The linked data field aims to integrate data sets, typically through more lightweight ontologies and direct linking of datasets via incomplete semantics. This emerged as a rethinking of the semantic web vision circa 2006 [350], through the 'Linked Data principles' [351]. The main usage of these principles has been in the field of linked open data, which aims to use linked data technologies to publish a single, global, open data cloud. Linked Open Data allows the sharing of public city data in a more powerful manner than raw data, and so greatly supports smart application development [352].

One significant difference between semantic web and linked data is the emphasis on the use of OWL, which has led to issues in reasoning over Linked Data, as combining RDFS and OWL graphs causes significant theoretical and practical problems [353]. This is very relevant to the discourse, as semantic inference is a key benefit of semantic technologies. Another significant feature of the linked data field is the use of the JSON-LD data format, which is an extension of JSON to facilitate more lightweight exchange of RDF data. This could also be used in a pure semantic web application as a serialisation format, which emphasises the overlap between the two fields.

It has been stated that the semantic web field is top-down, whilst the linked open data field in bottom-up [302]. This means to express that the semantic web aims to fit data within formal domain conceptualisations, whilst linked data adds minimal semantics to the data, as necessary for specific use cases. This minimal ontological commitment makes the Linked Data approach more palatable in the absence of accepted ontologies in a domain, or the absence of uses cases for comprehensive ontologies in a domain. This allows wider audiences to engage with semantic technologies, which is at the core of the internet's popularity and success. Given the absence of comprehensive ontologies for many smart domains, this justifies the linked data approach adopted by many smart city initiatives.

As both linked data and semantic web utilise W3C standards and address semantic aspects of interoperability, they are occasionally perceived as opposing technologies, although they are technologically highly compatible. In this way, they can be regarded as adjacent alternatives in the spectrum of semantic technologies.

### 2.3.6 SEMANTIC TECHNOLOGIES IN THE ERA OF BIG DATA

Big data is one of the most prominent and transformative trends of modern times. Big data is related to IoT in the discourse, and both emphasise a shift in paradigm due to unprecedented scale, rather than a revolutionary technology. This scale however, is only one aspect of the 'big' of big data, which is often expressed in terms of *volume*, *velocity*, and *variety*, and also *value* and *veracity* [350]. In this way, vast heterogeneity must be supported whilst prioritising response times. Semantic technologies are well equipped to support the variety, veracity, and value of big data, but there are concerns regarding volume and velocity [354], as exchanging the meaning of data alongside the data itself inherently increases message size. Therefore, leveraging semantic technologies in such a landscape requires a well-considered approach to scalability.

Literature broadly integrates the concepts of semantics and big data in two ways: i) developing a more efficient format for storing and exchanging semantic data, or ii) strategically designing a system architecture. Regarding the first approach, semantic message formats for constrained IoT environments [25], [327]–[329] are very relevant here, as these mitigate 'semantic overhead'. However, a more scalable way to manage the data is also needed. Map-reduce techniques are one promising option, such as through the Apache Hadoop framework [355]–[358]. However, RDF data stream overloading and reasoning remain significant open challenges in this area, due to the velocity and ordered nature of the data Aufaure et al. [35].

One option for RDF stream processing is to reduce the size of the ontology reasoned over through intelligent splitting, or using a lightweight ontology for the bulk of the inference, and having dedicated services for outlier tasks [354]. One example is the Waves platform, which used Apache Storm, Apache Kafka, and Redis, to handle IoT data following conversion to a compressed RDF format [359], through continuous SPARQL queries. A Kafka-Spark architecture has also been used for high velocity stream processing [360]. One aspect of the big semantic data issue is that more expressiveness typically results in more inference time; which requires research on fast reasoners [354]. As well as pursuing modular ontologies, incremental and distributed inference is a promising avenue to this end [315].

As well as the semantic web community seeking relevance in an era of big data, the big data community is also turning to semantic technologies organically to solve interoperability issues. One example of this is federating Cassandra and Solr databases as SPARQL endpoints [361]. Another example from the big data community considered how existing Extract-Transform-Load frameworks are insufficient for big data interoperability challenges, and explored the value which semantic technologies offer [362], [363]. This is corroborated by other work which states that entity-relationship and UML approaches are insufficient for ETL processes in data warehouse environments [364].

Some recommendations for further work to converge the semantic web and big data communities and technologies include i) requirements engineering work, ii) better ontology repositories and metadata to support reuse, and iii) better tooling for modularity as well as ontology engineering [354]. It is clear from the literature that this is a very recently observed challenge with significant research still required, but many promising avenues and early work towards addressing issues. It is important to establish the difference between using semantics to assist with developing interoperable applications and using 'big semantic data' in runtime applications. This is discussed further at the end of this chapter.

### 2.3.7 SMART CITY ONTOLOGIES

In order to realise the value of semantic technologies in smart cities, there must be sufficient ontological representations of smart cities and their sub-domains. Currently, this is a significant gap in the literature: the ISO/IEC Joint Technical Committee's report on smart cities [365] highlighted this need for ontologies. However, some work in this space is observed in the literature. For example, Fonseca et al. discussed issues related to urban ontologies for geospatial purposes [366] in 2000. They observed a number of complexities, including that most boundaries in cities are abstractly defined, most objects are complex, and very little knowledge exists outside of human perception.

IBM developed the SCRIBE smart city ontology 5 years ago [367], commenting on a lack of available ontologies, and stable OWL tools. Their ontology paved the way for formal descriptions of city services, events, metadata, and abstractions. A smart city ontology termed 'Knowledge Model 4 City' was developed in [368], with a focus on public transport and mobility, but including a mapping to sensors concepts. This simple ontology seemed only to facilitate the query of public transport data by SPARQL. The SEMANCO project developed a large smart city ontology in OWL DL-Lite$_A$ for the purpose of data integration [369], resulting in 592 classes. The SEMANCO ontology appears to be intended for the exchange of static data in the planning phase of urban areas though, given the lack of sensor concepts and dynamic data provision. It could therefore contribute to an upper ontology which links ontologies for each vertical which address these issues, as operational data and semantics are highly specific and 'owned' by each industry. The CityGML standard [370] formalises concepts and relationships relevant to geospatial knowledge in cities, and some semantics as to the nature of objects and spaces in cities, but in an insufficient manner for interoperability of operational smart city data across verticals. The ISO 37120 ontology defined indicators for sustainable communities [371]. A Linked Open Data modelling approach to managing smart city data is proposed by Consoli et al. [372], who federate GIS data into a lightweight ontology. Finally, BSI:PAS 182 [16] proposes a high level smart city ontology, which serves as an important step, albeit a 'lowest common denominator' approach, which hence captures little semantic depth. These efforts represent steps in the right direction, but the majority of the smart city domain remains to be modelled.

The SSN ontology which describes sensing devices, sensor networks, and the observations from these, has been broadly adopted, and should be leveraged wherever semantics and intelligent sensing are combined. The reuse, extension and alignment of these existing works will allow great extensibility, without vendor lock-in, and would allow knowledge re-use for future applications. The following section presents the cloud platform developed, before its semantic modelling, use cases, and validation are discussed further

### 2.3.8 SEMANTICS IN SEMANTICS: WHAT IS NOT AN ONTOLOGY?

As the definition of ontologies is contested and intentionally abstract, there is much conflation of related terms, and misuse of the term 'ontology' in practice. Based on the literature, a 6-level classification scheme of ontological resources is proposed in Figure 12, based on describing the water utility domain as an arbitrary example. This shows simple artefacts such as controlled vocabularies with minimal semantic formalisms up to full ontologies which are computable, and a further level of complexity can also be included which results in the ontology's computability not being guaranteed in finite time. These levels are now briefly described:

1. A controlled vocabulary defines the words which can be used in a domain, but doesn't specify semantics beyond these.
2. A dictionary proposes a controlled vocabulary and also offers human readable descriptions of these words, which aids in building semantic interoperability.
3. A taxonomy extends this by offering machine-readable classifications of objects, through 'type of' and mereological relationships, to produce a tree-like structure, as often seen in biological classifications.
4. A simple ontology extends a taxonomy by enabling further relationships to be machine-interpretable beyond 'type of' relationships, to express more complex aspects of a domain.
5. A computable ontology then extends this further by formalising greater semantic depth, such as restrictions on language usage, logical rules inherent in the language, and cardinality restraints.
6. Finally, a full ontology, which has few direct uses in semantic web software due to computation time, removes all restraints on domain statements to include situations such as asserting that a class is also an individual.

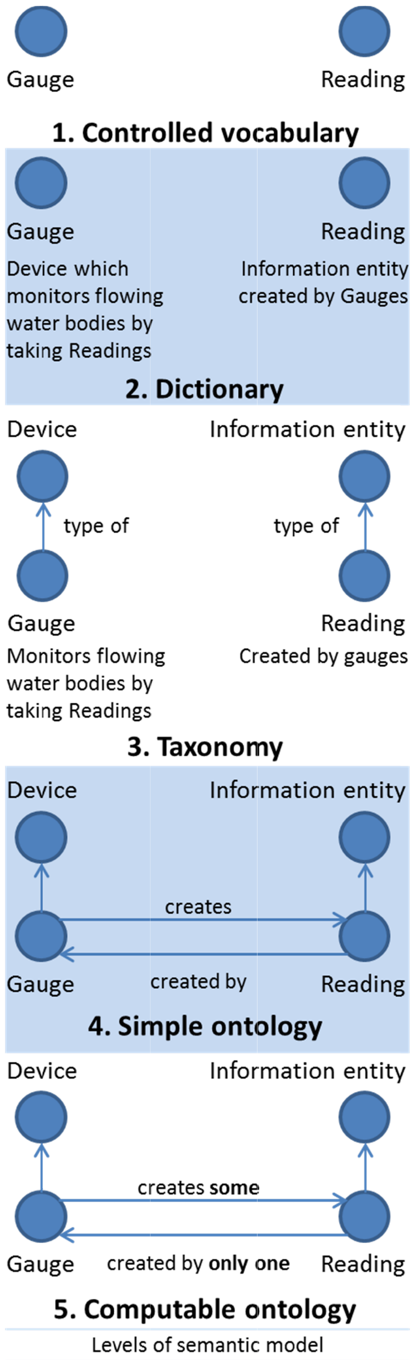


**Figure 13: Levels of ontological resource complexity**

### 2.3.9 ONTOLOGY ENGINEERING

Ontology engineering methodologies can be broadly categorised into manual, automatic, or semi-automatic. In general, manual approaches take significantly more human development time, but produce a model more closely coupled with a target system [373]. For either manual or automatic ontology development, the initial scoping and requirements engineering stage is of pivotal importance and typically ends with a set of competency questions. This is discussed briefly before manual and automated approaches are explained.

Competency questions provide a measurable objective which, once satisfied, indicates that the ontology's basic structure is appropriate and that a sufficient level of detail has been achieved [50], [308], where an ontology should only contain enough detail to meet the requirements. This highlights the importance of effective requirements engineering. A leading framework for ontology engineering is the NeOn methodology [313], which comprehensively describes ontology engineering activities and paths. The approach emphasises the early stages of knowledge gathering, feasibility studying, and requirement specification.

The development of ontologies for semantic web of things applications presents unique challenges, compared to those developed before the growth of IoT & SOA, which must be considered when evaluating more traditional methodologies. Whilst many of the recommendations and best practices of approaches such as Uschold [51], [374], [375] still stand, specific activities are best guided by recent works. METHONTOLOGY [312] has been well regarded for some time [376], but predated NeOn significantly, which itself is now 8 years old, although is still active. A more recent work [308] specifically addresses the development of ontologies for the semantic web, and incorporates many familiar themes, but still describes the field as immature. [308] also calls for a balance between object oriented programming model development and open-world ontologies.

Regarding automated ontology generation, the main stages once a corpus has been established are concept extraction, taxonomy extraction, and non-taxonomical extraction [377]. Liu et al. [378] describe existing methods and systems towards ontology extraction from free-text documents within the biomedical field. They outline exactly the same steps in the process as Chen and Williams [379], and group approaches through 'symbolic' and 'statistical' categories, although they favour research within the biomedical field. Several methods of automated ontology extraction have been proposed in the literature, such as compound term heuristics and Naïve-Bayes semantic labelling [377], statistical triple-based identification of noun-verb-noun triples [380], WordNet based sense disambiguation [381], or a simpler lexico-syntactic pattern identification [382].

One example [380] proposes a methodology for extracting complex relationships between concepts. First, assuming an existing concept list is available, concept pairs are extracted consisting of a subject and an object. Then, the authors extract candidate labels (verbs) by assigning a metric to all the verbs present which favours verbs which occur with a small set of concepts, to remove generic verbs such as 'do' or 'have'. Candidate triples are then extracted which consist of a concept pair from the extracted list and a verb from the extracted list. These triples are then assigned a metric based on the probability of them occurring by chance, and a probability threshold is used to filter less important relations.

This section has discussed existing progress and issues related to the integration of semantic technologies with IoT technologies and smart cities. However, the adoption of smart technologies and IoT must occur not only in new smart city departments, but across existing industries and organisations. The following sections discuss progress and challenges in smart city domains identified previously as pertinent: smart buildings, smart energy systems, and smart water networks.

## 2.4 SMART GRID: DELIVERING THE POTENTIAL OF THE EMERGING ENERGY LANDSCAPE

As one of the primary domains used to test the value of SWoT technologies, a thorough literature review was conducted of the energy domain around the relevant topics.

The concept of a smart grid has been defined normatively as "the integration of power, communications, and information technologies for an improved electric power infrastructure serving loads while providing for an ongoing evolution of end-use applications" [15]. Some of the main requirements of this have been defined [383]: self-diagnostics, optimization capabilities, topological adaptability, adaptive protection, distributed management, islanding modes, ancillary service provision, demand side management, improved forecasting, self-healing capabilities and preventative maintenance. Other mentioned requirements include consumer focus [63], [384], bidirectional data and energy flow [63], market efficiency and integration [385] and higher quality of service [385].

This section therefore begins by outlining the evolution of the energy landscape over time, before discussing the smart management of distributed energy resources, and then analysing existing semantic models of this domain and the relevance of multi-agent systems.

### 2.4.1 A BRIEF HISTORY OF ENERGY SYSTEMS

Following the first generators of Michael Faraday, electrification began using renewable sources, with the first public electricity supply based on a water wheel in 1881 [386]. The era of centralized power stations then began circa 1890 with the completion of the first high-voltage AC coal power station [387]. Regional and national grids grew in the 20th century, and due to economies of scale and inexpensive fossil fuels they became the status quo in the middle of the 20th Century. These began as public resources, but after the US Public Utility Regulatory Policies Act (PURPA) of 1978 [388] and the UK's privatization of other sectors, the UK deregulated and privatized its energy market in 1990 [389]. Hydroelectricity also has a prominent place in history due to its reliability, controllability, dispatch-ability, and storability, but other renewables traditionally had a negligible role.

Centralized electricity systems consist of a small number of large power plants, a high voltage transmission network, transformers which reduce the voltage, and medium and low voltage grids. For example in the US in 2005: 3618 plants (coal, nuclear and gas) supplied ~90% of the demand of 110 million households and 5 million commercial buildings [390]. In this approach power flows in one direction, and consumers are passive. This lack of predictability results in the need to store a reserve capacity in case of load spikes [391], which wastes energy. Also energy losses are incurred over the long transmission distances, and the heat by-product is wasted. Despite concerns, sustainability was only considered to be a 5-15 year objective [392] circa 1995, and a common view circa 2001 was that renewables held untapped potential [393].

Academic interest in renewables accelerated circa 2000, but inertia in the industry prevailed until circa 2005, when renewable reliance accelerated; doubling between 2004 and 2014 [394]. A number of pivotal 'top-down' actions such as Agenda 21 [395] and the Kyoto Protocol [396], officially recognized climate change at an international level and put steps in place to mitigate human contribution to it whilst adapting to its effects. National policies have since promoted and incentivised renewables, which has mobilized the energy landscape to change [394]. Whilst coal and oil reliance has been reducing, reliance on gas has been increasing alongside renewables due to its low emissions, and its alignment with existing markets and business models. However, natural gas has been predicted to deplete circa 2064 [397], so renewable and nuclear energy sources remain important.

In the initial stages of renewable adoption, energy systems have remained centralized by incorporating large wind farms [398], solar farms [399], and geothermal and biofuel plants, which have gradually supplemented conventional energy sources [394]. However, research has demonstrated the value of integrating distributed energy resources (DERs) into the energy landscape through distributed generation, polygeneration, active consumers, energy storage, plug-in vehicles and virtual energy management. This broad field of research has converged through the term ‘smart grid’ [400], which aims to intelligently manage new approaches such as microgrids and virtual power plants [63]. The main reason for a distributed grid emerging was the push towards resilience and sustainability, and advances in DER research which enabled new entrants to the electricity market [66], [87], [401], [402].

### 2.4.2 SUPPLY-SIDE ENERGY MANAGEMENT INNOVATION

This section describes the value proposition of the smart grid vision being pursued by research, which is centred on the intelligent management of distributed generation, polygeneration, energy storage, plugin electric vehicles, and active consumers.

A central feature of the shifting energy landscape is distributed generation (DG). This refers to the production of useful energy near or at the location of its use, which reduces transmission losses, and includes plants which power districts as well as even smaller microgeneration units. Wind and solar plants are common options, as they are more flexible in location than small hydro and geothermal, don’t incur emissions after installation, don’t require biomass fuel to be delivered, and are economically competitive. However, these sources of energy are subject to stochastic weather variations, so a core research goal has been mitigating this whilst ensuring quality of service to consumers and using renewables maximally and economically.

A common research avenue is to present an aggregated connection to the grid from a microgrid [403], [404] or virtual power plant [72]; which often include an energy storage solution [405] to store excess renewable energy. Microgrids are independently controlled distribution networks capable of operating in island mode [406]. The efficacy of islanding has also been shown on a real system in Hachinohe [403] and on a microgrid around a hospital [407]. Optimising a microgrid (with PVs, wind turbines and batteries) can reduce weather-based variation and effectively reduce CO2 emissions by 70% in real-world conditions [403]. These environmental benefits can be formally prioritised, such as through fuzzy logic and a multi-objective genetic algorithm [408]. Also, the profit of aggregated units can be optimised through a unit commitment approach in a day-ahead bidding [31], [409], or day ahead scheduling can be coupled with intra-day optimisations and topological reconfigurations [410]. These reconfigurations can also be used to minimise outage time [89]. The supervisory optimisation of simple systems can be achieved through rule based control [404], [405]. More complex examples benefit from machine intelligence such as neural networks [411], and the most promising research adopts a distributed intelligence approach [412]–[415] for greater extensibility, adaptability, and resilience.

Virtual power plants are entities which act between the grid and a collection of DERs to improve their operational characteristics through aggregation, and consist of dispatchable generators, stochastic generators, active loads and energy storage systems [72], and sometimes plug-in vehicles [73]. Whereas microgrids consist of a number of proximal and physically connected DERs and loads, this is not needed for virtual power plants, which aggregating them remotely as a market entity [416]. VPP management has been reduced to an economic optimization which considers the grid as an energy sink in the day-ahead market [72], where uncertainty can also be accounted for [409]. However, VPPs can meet technological objectives, such as load balancing, as well as economic ones [73], and can act across multiple energy vectors, such as through micro-CHP clusters [417]–[419].

The main viable forms of energy storage have been stated as pumped hydroelectric, battery storage and superconducting magnetic energy storage systems [420], although others are observed in the literature. Batteries are a common feature of microgrids, where they can reduce fluctuations from weather-dependent renewables, maximize renewable contribution to the local generation mix and regulate voltage and frequency [421], especially as they have dramatically reduced in cost in recent years. The ROI of batteries can be improved by coordinating short and medium term storage capacities to maximise their lifespan, without affecting the battery's response time [411]. Long term energy storage is also possible through the production and storage of hydrogen via electrolysis [422]. Another novel means of storage is the use of a district heating network as a heat sink [423].

As well as reducing transmission losses through distributed generation, recent technological advances allow waste heat to be reused through polygeneration, for example to contribute to a district heating network, cooling network, or even hydrogen storage. This can integrate the management of local heat and electricity networks, and the national grid [78], to create a multi-energy system [74]–[79]. This has shown operational cost savings of 20% [75], also allows flexibility to meet peak demand through overproduction at off-peak times [423] and the storage of energy to sell to the national grid at times of highest price, where dynamic pricing occurs [424]. The main challenges are the investment and operational costs [425], and the complexity of the multiple energy vectors and agents, especially if there are several micro-CHP plants [426]. Research has worked to minimise these costs, such as through dynamic programming based on graph theory and a black box models [50], and in more complex scenarios through multi-agent systems [83].

Extending the polygeneration concept further, energy hubs are grid nodes with multiple input energy carriers as well as output carriers. This allows a close integration of different energy systems through energy generation, storage, and conversion units at a single point in the network, which promotes increased reliability, load flexibility and efficiency gains [427], and significant potential for optimisation [427]–[429]. Again, this optimisation can be extended beyond financial objectives to include environmental concerns, such as the “social cost” of $CO_2$ emissions [430].

### 2.4.3 DEMAND-SIDE ENERGY MANAGEMENT INNOVATIONS

#### 2.4.3.1 TOP-DOWN APPROACH: SMART METERING, ACTIVE LOADS AND SCHEMES

Demand side management (DSM) refers to the systemic interaction with consumers and active loads to directly or indirectly affect demand profiles. This typically aims to either reduce the total load on the grid, the peak load on the grid, or both. This ability stems from the key feature of smart metering in smart grids: digital monitoring and regular transmission of consumption data to the energy provider, sometimes with bidirectional communication [431]. The concept of an ‘active load’ extends this to directly controlling aspects of the consumption. These represent transformative opportunities for managing energy grids, but also entirely new aspects of complexity in terms of security, privacy, trust, and psycho-social behaviour. This also opens up new market opportunities for ‘demand response providers’ as intermediaries between consumers and suppliers, with positive effects for grid operators [432]. Multi-agent systems are well suited to represent this distributed agency and complexity [83]. One central concept in DSM is dynamic pricing, which could be used to optimise across local and global goals through automated load scheduling [433], resulting in a more stable demand profile and lower overall system costs, which could then be passed on to consumers. In order to ensure a balance between local and global optimisation, distributed intelligence is fundamental [81].

As well as DSM of buildings, the growth of the plug-in electric vehicle (PEV) has led to unique challenges and opportunities, as they represent a significant total quantity of load and potential storage within the system [434]–[437]. They also exhibit stochastic demand profiles if not managed intelligently, which can cause large demand spikes at already peak times. Research has responded to this through intelligent scheduling [438], using PEVs as energy storage [439], optimising their market interactions, and intelligently managing large charging stations [434]. The ability of vehicles to sell energy back to the grid raises opportunities and complexity, and introduces new socio-economic aspects, although this uncertainty can be accounted for within model predictive control approaches [440]. The concept of aggregation can also be applied to electric vehicles, in order to improve their economic viability, and impact on the grid.

#### 2.4.3.2 BOTTOM UP APPROACH: CHALLENGES AND OPPORTUNITIES IN BUILDING MANAGEMENT

As well as optimising a grid's demand profile through incentives or schemes from the supplier, energy-saving retrofit measures and intelligent management can greatly reduce the energy consumption of buildings. The relevant aspect of this is the integration of buildings management systems (BMS) and building automation systems (BAS) with innovative IoT solutions and advanced applications such as artificial intelligence. A key facilitating field for this innovation is the use of Building Information Modelling, as this serves as a well-established set of processes and technologies for delivering data which is open and standardised.

Technology within building management typically adopts a familiar stack, with industrial systems, sensing and communication hardware sending data to a centralised location and acting on commands, either as part of a local feedback loop such as with PID, or as part of a centralised supervisory control process. Where a centralised monitoring point exists, it is often possible to access this data remotely, allowing the development of value-added services, and integration with more advanced applications. This stack has been described as three layers: sensor, computation, and application [441], although a middleware layer has also been proposed between the sensor and analytics layers [442]. This middleware layer plays a critical role in supporting the intelligence applied at the application layer, and more advanced applications require more comprehensive support from middleware regarding data meaning. A lack of sensing infrastructure is commonly found in existing buildings, which favours the use of simulation [443], [444], and surrogate models [445], and hence further emphasises the need for powerful middleware in retrofit buildings.

A great deal of research has focused on the intelligence of building management, such as through rule mining [446], neural network [447], fuzzy logic [448], genetic algorithm [449], ant colony optimization [450], and hybrid algorithms [451]. Machine learning has been demonstrated for predicting the behaviour of these highly nonlinear problems [452]. Either historical data or simulation-based data can be used to support artificial intelligence at the application layer of building management. For example, fuzzy controllers can be tuned by genetic algorithms over historical data [453] and data mining can produce integrated rule sets [446]. Recent simulation-based approaches are more effective, such as the use of EnergyPlus within a model-predictive control process [454], or embedding a simulation within a stochastic search method [90]. These two strategies can also be brought together into a hybrid approach, which offers several benefits [455].

Despite research offering promising avenues for promoting intelligence in building management, these efforts are typically tested *in vitro*, where the lower layers of the technology stack are assumed to be sufficient, or are developed ad-hoc. This has prevented such approaches being adopted in commercial BEMS systems. For example, MonaVisa is a retrofit solution which uses simple rules to notify a decision maker when a KPI leaves a set range [456], and PlugWise uses IoT technology to monitor temperature, motion, and energy consumption, but again only provides simple dashboard-type intelligence at the application layer [457].

One research field where solutions to the application-layer interoperability challenge may arise is Building Information Modelling. This aims to produce digital models and standardised levels and formats of data. For example, the Industry Foundation Classes (IFC) [248] are an ISO adopted standard for exchanging building data between software in the architecture, engineering and construction fields, with a primary emphasis on the construction phase of a building's lifecycle. This provides a comprehensive set of concepts for describing geospatial and semantic aspects of buildings. However, the model is currently based on the STEP-EXPRESS language, which has drawbacks, as discussed later.

### 2.4.4 EMERGING IMPORTANCE OF INTEROPERABLE, DISTRIBUTED AND INTELLIGENT SYSTEMS

Given the growing options for smart interventions in the energy domain at both the supply and demand sides, the impact of this must be considered, and the challenges in unlocking value from this should be pre-empted. Most research in this area has investigated the integration of a single new technology within the current centralised energy system, although it is clear that if DER penetration increases, the challenge will instead be how best to leverage a landscape rich in DER technologies. As the density of DERs and DER management structures increases, the potential benefit from coordination across these structures as well as the challenges associated with their integration with the grid increase dramatically. Some of the opportunities include flexibility, modularity, redundancy, coordination, aggregation, new market opportunities, systemic management schemes and big data, as well as the sustainability benefits in increasing RES reliance In order to achieve the most impact from this scenario, it is pertinent to research and take measures now to promote interoperability such that new control strategies can be implemented as the number of 'nearby DERs' gradually increases.

Consideration of this in the literature is very sparse, but some authors have begun to consider interactions between management structures [71], [91]. In order to leverage the system of systems nature of energy systems to improve their operational performance, research must progress towards solutions which are truly scalable, utilise artificial intelligence maximally, and prioritise interoperability to cope with unavoidable heterogeneity. Further, given the stochastic nature of many of these local energy solutions due to their reliance on RESs and human behaviour, dynamic management of the network must be based to a greater extent on probabilistic simulations and shorter time-scale reactive management [84]. Integrating data heterogeneity is one of the main purposes of semantic technologies, and advanced multi-agent system research shows promise at coping with the required scalability. For these reasons, the following section presents recent progress towards distributed intelligence in energy systems, and then the progress of semantic technologies in this area is presented.

### 2.4.5 TOWARDS DISTRIBUTED INTELLIGENCE

Agent-based technologies and distributed intelligence are promising and active avenues of research in this area [458]. An agent is a software entity which exhibits autonomy and goals. The behaviours of agents are conditioned by their individual goals, and can cooperate or compete with other agents. The behaviour of the overall system then emerges as a result of its agents. By designing the agents, and their interactions and goals carefully, this emergence can be used to optimize the system [240]. Also, the intelligence becomes more adaptable, resilient and scalable than centralized approaches [31], [239], [243], [459]. This resilience is due to each agent responding to changes in the system's structure and components automatically, which tolerates partial failures in the system and leads to adaptability [31], [88], [91], [241]–[245].

The main approach observed in the literature is to develop device (and possibly supervisory) agents, then to simulate their efficacy for the authors' intended purpose in coordinating electricity supply in an example network [67], [412], [414], [415], [460], [461]. One of the early seminal market-based efforts was the PowerMatcher solution [461], [462]; a supply and demand matching system which aimed to promote sustainability in urban energy systems. Over the past decade, PowerMatcher has been developed and validated in real-world settings [400], [462], [463], and extended to consider electricity storage and microgrids [91], [412], [464], integrated heat and electricity systems [241], and DHN systems [465]. Dynamic pricing and market-based approaches are now well explored by various authors [91], [242], [412], [414], [464], although most examples typically exhibit a simple hierarchical structure, and only undergo lab or simulated validation. Van Dam et al. managed a collection of micro-CHP units in a VPP through agent based control [418]; which was then incorporated into the PowerMatcher repertoire [466]. Also, agent-based approaches have been explored to enable more accurate set point scheduling through demand forecasting [67], [467], and Lagorse *et al*. improved system resilience [459] by using a virtual 'token' to decide which device agent is responsible for ensuring the bus voltage.

Urban energy management is an increasingly complex and multidisciplinary effort, with vast heterogeneity between components; causing significant interoperability issues [15]. It is critical that intelligent entities share an understanding of the domain, such as through common vocabularies, data models [92], [97] and ontologies. This is especially important within multi-agent systems [243], where FIPA-ACL ontologies offer a step in the right direction as an enabling technology, albeit lacking the expressiveness of OWL.

### 2.4.6 PROGRESS AND CHALLENGES IN APPLYING SEMANTIC TECHNOLOGIES

Semantic technologies are playing an increasing role in energy systems in supporting more powerful interoperability, and artificial intelligence functions. Interoperability is increasingly noted as a critical concern [15], [92], [93], including semantic heterogeneity, interoperable protocols, data formats, data quality, security, and trust [15], [16]. Semantic technologies can be used to reduce the development effort of advanced applications which use distributed data sources [15], [468]. This section briefly outlines the standards in this space, before discussing ontologies and the applications observed in research.

Semantic energy standards typically adopt an entity-relationship approach rather than a triple based approach, and generally lack both expressivity and consideration of emerging technologies in smart grids. The main industrial standard is the Common Information Model of IEC [469], but this doesn't support the increasingly distributed nature of energy systems. Notable standards which aim to overcome these challenges include the openADR model [470], and the energy@home data model [471], and from the BIM domain the IFC [248], [472] and SAREF [473] models are relevant, all of which are now discussed.

The International Electrotechnical Commission (IEC) has identified over 100 standards which it deems relevant to smart grids, of which the Common Information Model (CIM), IEC 61970 and IEC 61968, form the core semantic model. The CIM was developed in UML [469], [474], and later bound to RDF [475], although it still lacks the expressivity offered by OWL [250]. The model is split into three layers and covers a broad spectrum of exchange cases including network management, compliance checking, customer billing, and risk planning. However, this model has some limitations and contradictions [250]. As the model is over a decade old, it is rooted in a paradigm of centralized energy generation, although research has attempted to incorporate modern concepts [476].

The Open Automated Demand Response (OpenADR) [470] specifications support smart grid concepts such as DERs and dynamic pricing through a data model and communication specification. These specifications model the middle concepts between consumer and supplier. Some of the main features are: opt-out capability, a rich data model, scalability, and openness. Decoupling the data model from the communication specification is a benefit, as this promotes reuse for other purposes.

The energy@home data model [471] covers smart grid concepts relevant at the domestic consumer level, and is broadly aligned with OpenADR. This describes smart appliances, power profiles, renewables, smart meters, and user interfaces. The model describes static device parameters, and also energy consumption profiles. These consumption profiles are detailed and hierarchical, whereby a profile consists of a sequence of modes and breaks, where each mode can be decomposed into phases and represents a function of the appliance, such as a specific washing machine cycle. The overall profile therefore combines numerous functions, for example representing a wash-dry program of a washing machine. Between the CIM, OpenADR, and energy@home standards, the entire spectrum of demand response is modelled from supplier to domestic consumer, and so integrative work in this space would be highly valuable.

The energy@home data model has similarities to the Smart Appliance Reference Ontology (SAREF) [473], which has emerged from the linked building data community, and has been standardised by ETSI. The SAREF ontology acts as a consensus smart appliance ontology, based on 23 analysed ontologies, and aims to unify them as a 'lowest common denominator'. This reduces alignment efforts in systems with 3 or more smart appliance ontologies. As well as specific energy ontologies, aspects of BIM are relevant to the energy sector, and the IFC have been extended for AI applications through object-based knowledge exchange [477]. The development of ifcOWL is an active area of research [472]. The semantic sensor network (SSN) ontology [330] described previously is also directly relevant here as an upper ontology which can be extended with descriptions of smart devices and the data they collect.

As an example of how ontologies can be applied in energy systems, the ISES project developed a BEMS which used an OWL-DL ontology for interoperability across lifecycle processes [478]. Also, the HESMOS project integrated data from building energy systems which were distributed and heterogeneous, through an ontology equipped framework [479]. However, these projects do not sufficiently reuse existing standards, such as the IFC, nor do they fully consider the social aspects of energy systems.

Most of the recent semantic modelling in the literature facilitated the planning or analysis of urban energy systems through simulation [480], [481] or information representation and exchange [249], [482], [483]. A few examples facilitated MAS communication [468], [484], [485] or complex event processing [97]. Unfortunately most semantic models intended for agent communication [484], [485] appear to be little more than class structures. The seminal work of van Dam [468] models urban energy systems as socio-technical systems in order to represent concepts such as ownership and contracts. This was compared to the SynCity ontology [483], by Keirstead and van Dam [486], who advocate that a common upper ontology would facilitate integration considerably even if not universally accepted. This is corroborated by Catterson *et al.* [487] and Zhou *et al*. [97]. The smart grid information model (SGIM) presented by Zhou *et al*. [97] is worth noting, and was designed to manage real-time sensor data in an event-based system.

The CityGML utility domain extension [488] extends the OGC CityGML standard to describe networks and object types, although unfinished and inactive for 4 years. Finally, the SEMANCO Energy Model [249] uses OWL to describe a vocabulary for energy infrastructure planning, and is based on graph theory. The CityGML Utility ADE and the SEMANCO Energy Model are broad models of city level data in terms of components and performance metrics, and are primarily suited for system planning and analysis rather than operational management. Key avenues for progress in this space would be integrative research, metamodeling, consensus building and adoption, more advanced use case driven modelling, and work closer to enterprise systems and *in vivo* testing rather than theoretical modelling.

## 2.5 SMART WATER: THE EMERGING IMPORTANCE OF SEMANTIC TECHNOLOGY

As the second main domain where the value of SWoT technologies was tested, a thorough review of the smart water domain was conducted, although this domain is significantly younger than smart grid, and so less literature was observed. Modern research, and the concept of a circular economy, emphasises a ‘whole value chain’ approach. A *water value chain* is defined here as all processes, agents and objects pertaining directly to the delivery of potable water to users and the subsequent use and removal of both foul and surface waters, from abstraction and treatment to usage, waste collection and disposal. This section briefly introduces smart approaches to managing this value chain, before discussing the role of interoperability and semantic technologies, and finally analysing relevant semantic models.

### 2.5.1 INTRODUCTION TO SMART WATER NETWORKS

The water sector is being transformed through the use of smart systems [107], [489] such as intelligent sensing [490], optimisation [491], and decision support [492], which aim to tackle sustainability and economic challenges. Smart water networks have been noted to promote efficacy, efficiency, and resilience in water infrastructure [489], [493]. Therefore, the amount of devices and software used is increasing rapidly, in line with broader IoT and AI predictions, and these resources must be used together efficiently.

A cluster of European Commission Seventh Framework Programme research projects, ICT4Water, has been formed to investigate this proposition, and the European Innovation Platform for water has launched an action group for water monitoring for decision support [494]. The smart water networks forum (SWAN) has proposed a framework [495] of a number of layers: physical, sensing and control, collection and communication, data management and display, and data fusion and analysis. Intelligent pressure management has recently been shown to reduce leakage by 12% based on an EPANET model [496], and can also minimise energy consumption [108]. Cloud-based machine learning has also been explored for leakage reduction [497], and a cyber-physical platform has been used to promote reliability, resilience, and energy reduction [498]. However, implementing these solutions in practice requires pervasive interoperability, as highlighted by the recent SWAN report on communication in smart water [112].

### 2.5.2 EMERGING ROLE OF SEMANTIC TECHNOLOGIES IN WATER

Applying IoT and AI in the water sector has much potential, as it does in the energy sector and other smart systems. This has been increasingly recognized across stakeholders over the past 5 years as a means to deliver water loss reduction, energy savings, water quality assurance, improved customer experience and operations optimization, amongst other KPI benefits [110], [499]–[501]. This is achieved through the use of advanced analytics to provide insight into complex systems through abundant data and an integrated approach.

However, the adoption of ICT in this field is hampered by the same interoperability challenges found in smart grids: (a) lack of machine communication protocols, (b) lack of common data formats and (c) lack of a common meaning of exchanged content [15]. This is corroborated by the ICT4Water cluster, who recently advocated for standard semantic models in this area [111], and aligns with the message of BSI regarding smart cities [16]. This leads to a clear emerging challenge; interoperability in terms of sharing meaningful data. Semantic models support this as a shared conceptualization of a domain, allowing contextualised data exchange. This allows management approaches which bridge traditional barriers within and across water organisations and technical systems. The role of IoT and semantics in smart water is beginning to emerge in research [502], but is embryonic at best. The following section presents notable efforts towards semantic models in this space.

### 2.5.3 TOWARDS A SUITABLE SEMANTIC WATER MODEL

The role of semantic technologies is increasingly recognised in the smart water field [111], [113], [365], [503]. However, little modelling has been conducted, widely adopted, or standardised in the water sector which meets the needs of smart water networks, so examples of smart water ontologies are sparse. Several mature ontologies were observed in the earth science field [504]–[507], but these are not suitable for the application of ICT to the water value chain, as they don't describe water utility networks. Also, the WaterML2 [508] and water data transfer format [509] standards are highly relevant to smart water, but do not express domain semantics. The main models observed in the literature are compared in Table 5 below.

**Table 5: Comparison of relevant existing water semantic models**

| Acronym/name | Description | Owner | #Entities | Date |
|---|---|---|---|---|
| SWIM | Device level IoT semantic model for the water industry. | Aquamatix | 41 | 2016 |
| INSPIRE Data Spec – Utility Network Model | The INSPIRE directive is establishing an infrastructure for spatial information exchange in Europe, resulting in data models for many application domains, including utility networks, of which water and sewer networks are | EC | 68 | 2013 |

| | | | | |
|---|---|---|---|---|
| | a subset. | | | |
| WatERP | Lightweight ontology of generic concepts for water sensing and management. | EURECAT | 25 classes | 2013 |
| Water Innovation Thesaurus | Aims to facilitate collaboration for water innovation by establishing and highlighting recognised terminology and providing clear definitions for these as well as demonstrating the relationships between terms. | EIP Water | 548 | 2013 |
| CityGML UtilityADE | CityGML ADE for the modelling of utility networks in 3d city models, based on topology and component descriptions. | OGC | 317 | 2012 |
| SWEET | Middle-level ontology for environmental terminology. | NASA | 6000 | 2011 |
| Hydrologic Ontology for Discovery | Supports the discovery of time-series hydrologic data collected at a fixed point. | CUAHSI | 4098 | 2010 |
| HydrOntology | Aims to integrate data sources regarding hydrographical information from a civil engineering or town planning perspective and a top down methodology. | Vilches-Blázquez et al. | 250 | 2009 |

The main relevant ontology observed was the WatERP "generic ontology for water supply distribution chain" [251], [510]. The semantic water interoperability model (SWIM) [511] is also very relevant, as well as the INSPIRE data model [512], and the CityGML utility network model [488]. SWIM formalizes a description of water sector devices such as sensors, pumps, reservoirs and valves.

The WatERP ontology only contains 25 classes, and few details of the physical processes and components involved in water management, and it doesn't describe relationships between features of interest or actors. The WatERP ontology is split conceptually into a 'supply and demand ontology', 'observation and measurement ontology' and an 'alerts and actions' ontology. Further, the WatERP ontology only captures high level concepts such as physical element types, and a few types of actors.

The INSPIRE Utility Network specification [512] formalizes simple concepts and relationships about water and sewer networks, but only includes 68 named entities. The CityGML Utility Network Application Domain Extension includes geospatial and semantic concepts about network entities such as pipes and manholes, and describes some properties of flowing water and pipe materials. Again, this is not comprehensive or semantically expressive enough, and has been inactive and incomplete for several years.

More mature efforts exist in neighbouring fields such as BIM, smart grid and environmental science, but these must converge with water industry standards such as WaterML2, whilst leveraging emerging models such as SWIM, and covering sufficient breadth and depth of water sector concepts. An integrated and standardised framework of these built environment, Semantic Web, and water industry standards would be a significant benefit for the interoperation of smart devices, repositories and artificial intelligence in the water sector. There is therefore a significant gap of capturing in-depth knowledge regarding the technological, network, social, sensory and ICT artefacts involved in water management decisions in a water value chain.

## 2.6 DISCUSSION: NEED FOR INTEGRATIVE RESEARCH IN THIS SPACE

The smart city paradigm can potentially deliver a great deal of economic, social, and environmental value, by improving the way a city's system of systems operates, by better informing all of its decision makers. However, work to date has focused on digitising rather than value-delivery, where data is centralised and displayed on a dashboard or visualised historically, but not used to derive deep systemic insight or business intelligence. This requires more advanced applications to be built on smart city data, a great deal of which resides in private organisations. Whilst the 'smart city' domain is conceptually composed of sub-domains, in reality 'smart city' initiatives are publicly funded and do not integrate these sub-domains' systems in a meaningful way. This is partly due to i) the emphasis on open data, which cannot include sensitive data about critical infrastructure or private businesses, ii) a lack of research on integrating enterprise systems with novel IoT solutions, iii) resources across domains being semantically heterogeneous, and iv) insufficient support for the integration of smart city data with advanced applications. Smart city platforms typically have not considered data semantics, but the latest and most advanced work is beginning to recognise the importance of semantic technologies [513].

It has been argued throughout this section that IoT is the emerging 'canvas' of smart cities, through which intelligent management must be applied. IoT has overcome several interoperability challenges, but whilst 'things' can now communicate, and securely discover and share resources [18], [47], [514], this has only fed lower-value knowledge delivery. In order for more advanced applications such as artificial intelligence and optimisation algorithms to use this data they must robustly and fully understand the meaning, provenance, and context of the data. This interoperability at the application layer has been coined by some as the ‘web of things’ [328]. By establishing application-layer interoperability through semantic modelling, the components of a smart city system could understand the meaning, context, and provenance of data. This would improve the value delivered by DSTs, through more powerful, reliable, and simpler integration of data and advanced software. To businesses this may represent reduce costs for future software development, reduced time to market, and less risk through a more systematic approach to dealing with complex webs of interoperation.

Several smart city IoT platforms have considered semantics, such as the ALMANAC IoT platform [18], which used a “semantic representation framework” to promote resource discovery. The SmartSantander project [515] also used a ‘resource directory’ as a repository for IoT resource semantics. However, these examples don’t couple IoT semantics with domain semantics and dynamic data, which is essential for genuine machine comprehension of data context for advanced applications. Further, it is clear that research is required on the potential of semantic technologies within enterprise systems, as most research has been on the open ‘semantic web’ [516], but much ‘smart city data’ resides within private organisations. There is a clear need for semantic models of smart domains, and integrative research alongside enterprise systems, IoT systems, and advanced applications.

Currently, domain semantics are dealt with implicitly and manually in building interoperating software systems. By proceeding in an ad-hoc, implicit manner, the time and cost needed to reach confidence in software will grow exponentially alongside the number of IoT resources, eventually becoming prohibitive as human comprehension of complex systems and semantics becomes limiting. This complexity is magnified in industrial systems, as is the need for confidence in the assumed semantics of data. Harbor Research stated that application integration standards in IoT was the biggest unmet need in 2012 [517], and this is still the case today. Explicitly defining semantics in ontologies could overcome this roadblock to unlocking the potential of smart systems.

The overarching argument is now formalised. The assumptions made are: i) machine intelligence has the potential to revolutionise the control of complex systems, ii) using machine intelligence in an operational context requires deep automated comprehension of consumed data, and iii) data comprehension can be supported through semantic technologies. It follows from these assertions that semantic technologies may support the use of IoT within advanced applications, which can offer significant value to decision makers in complex systems such as smart city domains. Through this chapter it has been observed that research around this use of semantic technologies is significantly lacking. The observed gaps in the literature are:

1. Elaborate cross-domain use case scenarios justifying the need for the smart cities paradigm across the city value-chain.
2. Requirements on semantic technologies for solving relevant challenges in smart cities, with a focus on these (point 1) scenarios.
3. Ontological representations of smart systems promoting inter-disciplinarity and cross-domain considerations.
4. Software designs, best practices and recommendations for leveraging semantic technologies alongside IoT and artificial intelligence.
5. Real-world implementations of the smart city concept demonstrating added-value to citizens and stakeholders across the complex city value-chain.

The thesis now continues by discussing the methodology followed whilst investigating the stated hypothesis, as an effort to contribute to filling these knowledge gaps identified in a rigorous and systematic manner.

# 3 RESEARCH DESIGN AND METHODOLOGY

## 3.1 INTRODUCTION

In order to take a step towards filling the gaps identified in the literature review through a rigorous scientific process, this chapter describes the research design and methodology adopted. Firstly, the philosophical and pragmatic paradigms in which the methodology is grounded are discussed. The research approach builds on established research design practices and modern developments to justify the processes undertaken to address the literature gap identified, and framed through the research questions posed in section 1.5. In this manner, this section links the literature review chapter with the findings chapters on which the discussion and contribution are based.

The justification and presentation of the research methodology in this chapter broadly follows the model proposed by Saunders et al. [518]: the interpretation of which is illustrated in Table 6 below.

**Table 6: Methodological aspects, adapted from [518]**

| Methodological Aspect | Examples, schools of thought |
|---|---|
| **Epistemology** | Objectivism, Constructivism, Foundationalism, Scepticism |
| **Research theory** | Positivism, Realism, Interpretivism, Pragmatism |
| **Approach** | Deductive, Inductive |
| **Strategy** | Experiment, Survey, Case Study, Action research, Participatory research, Archival research |
| **Choices** | Mono method, Mixed methods, Multi-method |

The chapter therefore begins by discussing the uppermost level of the identified model of research design; the philosophical stance adopted. The chapter then proceeds by discussing the high level research approach adopted, followed by the strategy and more detailed choices, before describing the pragmatic processes undertaken and the roles of these in contributing to answering the research questions. This entails a description of 3 distinct stages in the overall study; theoretical analysis, participatory action research learning cycles in the energy and water domains, and a design research process to unify and generalise learnings and artefacts from across the previous stages. Each of the identified stages' roles within the overall research design is described and each of their systematic methodologies discussed, before they are expanded on alongside their findings in the following chapters.

## 3.2 EPISTEMOLOGICAL AND PHILOSOPHICAL PERSPECTIVE

Epistemology is the philosophical study of knowledge, rationality and justification [519] and is arguably the most abstract perspective necessary from which to begin a discourse regarding a research design [520]. This is because epistemological choices define what the researcher believes constitutes valid knowledge and describes how they perceive existence. Following a consideration of the merits of concepts such as objectivism, constructivism, and subjectivism, these abstract notions can then begin a discussion of the theoretical underpinnings of the research approach, transitioning towards more direct relevance to the research design in a rigorous manner. Within epistemic justification, a central aspect of epistemology, some main schools of thought are foundationalism, coherence theory, reliabilism, scepticism, and cognitive realism [519], [521], [522]. Whilst a comprehensive and detailed consideration of such concepts is beyond the scope of this work, it is important to ground the research design within philosophy and theory, to validate the knowledge contribution of the work conducted and conclusions drawn. This theoretical grounding has been stated as the difference in the applied field of information systems between researchers and software developers or consultants [523]. This section now discusses the aspects of philosophy and research theory necessary to rigorously ground the knowledge contribution.

Saunders et al. state that the first aspect which should ground a research design is a research philosophy, which they describe as a high-level term regarding the development of knowledge and its nature [518], which broadly concurs with the aforementioned work of Gray [520]. The authors go on to explain that this choice fundamentally affects subsequent research strategy choices, and imports underlying assumptions into the research and hence the knowledge it produces, and that the choice is likely to be influenced by the researcher's perspective, and practical considerations. It has been argued that research philosophy and methods should not be viewed independently, and that research philosophy should be addressed pragmatically, mixing models as required to answer the specific research questions posed [524]. It follows then that enshrining a single philosophical doctrine with regards to research design may be detrimental to the varied insights which can be elicited from research practice.

The views of Saunders et al. regarding philosophical stances are broadly echoed by Ritchie and Lewis [525], who explain the role of philosophical paradigms as a foundation of research design. Both sources identify positivism as the most clearly defined and acknowledged stance, which holds that objective truth exists and should be gained through observation and empirical analysis, with the researcher remaining objective throughout. This contrasts post-positivism, which acknowledges a researcher's lack of neutrality and attempts to account for this in research design. These further contrast anti-positivism (referred to as interpretivism by Saunders et al.), which holds that objective truth should not be the goal of research, but that reality and truth are experienced and interpreted by individuals [526]. Interpretivism is described as a group of several schools of thought, including notably, constructivism and phenomenology [518], [526]. Constructivism, as an epistemological philosophy, argues that reality is affected by the research process, whereby truly objective research is not possible, and researchers can choose to remain neutral or to personally engage in a study [525]. It is highly relevant that Ritchie and Lewis state that natural science research methods are not appropriate for deriving knowledge about the social world [525]. This is pertinent given the critical social dimensions of smart cities, and of ICT solutions in both the deployment and development stages, whereby social and human-machine interactions are a central aspect of testing the hypothesis.

Despite the clearly separated schools of thought, Saunders et al. state that a consideration of philosophical paradigms should not be viewed as a 'shopping list' from which to choose the best option. Instead they argue that understanding the implications of philosophical stances on research design is valuable to enhance the production of valuable knowledge within a particular field [518]. To this end, and given the mixed social and technological aspects of the research questions considered in the current study, the paradigm of pragmatism holds significant value, and is again described by Saunders et al. This approach holds that choosing and enshrining a paradigm may not be ideal, and that the philosophical stance adopted should be tailored to the research questions, hence supporting the role of mixed methods research, engaging with both quantitative and qualitative data. This concurs with the much earlier work of Orlikowski and Baroudi [527], who argue that augmenting a traditional, positivist, research perspective with interpretivism would produce better research endeavours.

Critically, within the computer science field, the discourse regarding research design philosophy is markedly different to natural sciences where positivism is the ‘gold standard’, given the unique nature of computer sciences of studying phenomena within man-made machines, bridging the science and engineering fields [528], [529]. This provides further evidence that solely pursuing objective truth is not appropriate for evaluating the proposed hypothesis, as the role of social agents within the systems to be studied cannot be negated whilst drawing meaningful insight about the value of semantics and ICT within urban decision making. Despite this, the role of empirical observation and objective truth cannot be ignored when studying phenomena such as software speeds and objective functionality. Therefore the combined use of these, with a focus on applied theories rather than pure theory, as is typical of pragmatism, is the paradigm primarily aligned with for the purposes of this thesis overall. It has been argued that whilst pragmatism has a clear foundation in empiricism, it goes beyond this to fully acknowledge the mutual permeation of knowledge and action in reality [530], and hence primarily concerns the actions necessary for study. Therefore, the implications of this philosophical stance on the research is to adopt a considered approach which prioritises the justifications for actions on a cases by case basis towards the overall research goals and design of each stage, rather than aiming to enshrine a single pure philosophical perspective. This is particularly appropriate for the current study, given the multi-staged design adopted, where each stage prioritises different research processes and hence benefits from a different, albeit compatible, philosophical stance.

## 3.3 RESEARCH APPROACH AND STRATEGY

The choice was made to adopt a pragmatic research philosophy, incorporating aspects of positivism alongside the interpretivism often used outside of natural science. The stance was hence malleable as necessary to suit each specific research question. This set a foundation on which to consider the functional research approach and strategy, leading into concrete actions and research methods.

Saunders et al. [518] discuss 3 research approaches in the context of the field of business: deductive, inductive, and mixed approaches. They again promote a defensible rather than a dogmatic approach to research design choices, advocating the role of combining deductive and inductive approaches. Also within their model of research design are the layers of ‘strategy’ and ‘choices’, and pragmatism suggests that these not be viewed independently, but rather stem from the needs of the research questions posed. This advocacy for a hybrid or multi-methods approach has been echoed in IS research [520], [531] and can strengthen research by ‘triangulating data’ to point towards research contributions from multiple corroborating directions [532].

The nature and contrasts of deductive and inductive approaches have been well considered in the literature as the two leading research approaches, such as the well-considered and modern work of Soiferman [533]. Inductive research is often described as a bottom-up process, where specific observations evolve into generic models, whereas deductive research is a top-down process, which moves from generic models to data. Soiferman draws parallels between quantitative and deductive methods, and between qualitative and inductive methods. In this manner, deductive reasoning is often associated with the positivist research paradigm, as quantitative data serves positivist researchers as a measurable and objective truth [534]. In contrast, inductive reasoning is often associated with interpretivism, where qualitative data is used by researchers to induce generalisms about social or philosophical science fields. Whilst these two fields of scientific thought have been thought of as competing for some time, Soiferman highlights that there is a growing precedent to move beyond this mindset. As with the decision to adopt a pragmatism-oriented perspective, the benefits of both deductive and inductive approaches could be leveraged alongside one another in a well-constructed methodology. Further to this, the relative peculiarity of computer science in comparison to conventional research fields such as logic, natural sciences or social sciences, renders the decision to dogmatically follow a long-standing epistemological viewpoint somewhat moot, as each field typically favours a philosophical stance, and computer science spans many of these conventional fields.

One branch of research strategy deemed particularly relevant for the purpose required is that of participatory action research, which has been advocated for in this field [535]. One clear definition of action research is offered by Herr and Anderson [536]: “action research is inquiry that is done *by* or *with* insiders to an organization or community”. Action research aims to balance immediate problem solving with knowledge generation [537], and has been an established practice in social and medical science for a significant period of time, and has also been advocated for in ICT since the 1990s [538]. This presumes that systems with complex social interactions cannot be reduced and studied as simply the sum of their parts. Further, the action research paradigm is stated to hold 3 prerequisites; an interpretivist viewpoint, an idiographic viewpoint, and the use of qualitative methods [538]. However, as previously argued, the present work does not dismiss the value of empirical analysis of some aspects of ICT systems, such as those which purely involve computational speeds without requiring a consideration of human-machine interaction. Participatory action research is then a subset of this type of systematic enquiry, and has been described as an extension of this wherein the ‘client’ also pursues information and ideas in a synergistic manner to the researcher [539]. This is particularly relevant to the field of IoT and urban cybernetics as it intrinsically involves the analysis of socio-technical systems, wherein ICT systems act primarily to support human decision makers.

Within action research, the researcher does not act as a neutral observer, but actively engages in the research in taking action towards overcoming an immediate challenge, whilst generating knowledge. This research stance is contrary to conventional positivist science, and some typical criticisms are that the research is less controlled, has greater risk of failure, and is interpreted subjectively, which Kock believes can be addressed through a number of ‘antidotes’ [540]. Further difficulties are considered and addressed by Marshall and Salas [537], who reflect on best practices for balancing the problem solving and research objectives of the approach based on a project they undertook. They explain that individual perspectives on the problem will differ within the organisation, formal business systems will differ from those implemented, the capabilities of the researcher(s) must be thoroughly understood and acknowledged, and effective communication between parties is critical.

As well as action research, case studies, and empirical research, design research has been legitimised in the study of information systems, themselves being an inherently applied discipline [541]. Design research aims to formally study the processes of design, with the goal of making design more empirical in nature [542]. In this manner, information system design research typically creates a business or management-oriented ICT artefact which extends the state of the art, and formalises knowledge from this which is beneficial for future designs [542]. Clearly, this narrative highlights similarities to action research, where the goal is to overcome an immediate challenge (such as through artefact design), and simultaneously produce knowledge (such as guidance for future designs). This similarity is observed in works which refer to both fields of study alongside each other, such as [543], which considers the nature of problem formulation in information system research, and discusses the implications for both action research and design science research. Given this precedent, the use of aspects of both design science and participatory action research are used within the current research, with an emphasis on the aspects appropriate for each stage of the research.

Whilst the research will primarily follow the methodological traditions of action research and design research in information systems, given the pragmatic paradigm adopted, a consideration of positivist approaches must be offered in the context of each research question. In this manner, the role of empirical case studies and experiments must be acknowledged. Specifically, in determining the viability of enterprise software solutions, there are a number of empirical metrics which may be observed impartially, such as computation speeds over benchmark problems. These metrics can be measured independently of the organisational context in which the information system is implemented, and is relevant in most settings. Therefore, the approach utilised a mixed-methods approach, which drew on some aspects of more conventional positivist research approaches where appropriate. As became evident, this was particularly necessary to thoroughly investigate the scalability of semantically enabled solutions, as the perception was observed amongst experts that such solutions inherently involve verbose messaging formats and ‘contextual overhead’, which reduced performance. Investigating this aspect and similar aspects of the designed systems benefitted from a more empirical perspective in order to generalise about the value of the approach adopted in ICT systems.

This section has discussed the literature regarding research methodologies in ICT, information systems, and applied computer science. The advocacy for an approach which combines established practices and bridges traditional divides, when justified, and the unique nature of the human-cyber-physical questions posed, led to an iterative and mixed-strategy approach being adopted. This supported the validity of the resulting knowledge, by establishing multiple viewpoints for the conclusions drawn. Further, the strategy and methods chosen at each stage of the inherently iterative action research approach adopted, was chosen based on suitability for the stage, and research question, under consideration. This mandated a well-considered approach to managing the overarching research design, which propagated the philosophical and higher-level decisions taken, through the practical sub-processes undertaken, in a coherent and logically consistent manner for deriving well-grounded knowledge. This overarching research design is now described in detail in the following section, and each stage's detailed and pragmatic methodology is introduced herein, then expanded upon within their relevant subsequent chapters.

## 3.4 RESEARCH DESIGN

### 3.4.1 RESEARCH DESIGN BACKGROUND

This section describes the pragmatic approach to collecting evidence and subsequently analysing this towards answering the proposed research questions. Formally considering and specifying a research design aims to facilitate an improved strategy and research integrity. This must detail the methodology conducted and artefacts used in producing and collecting data for analysis. In order for the methodology to elicit the best data from the systems under consideration, Cooper and Schindler [544] describe the essential components of a research design:

- A plan which details both the actions and durations and chronology of the research
- A clear focus on specifically answering the research questions

- Decision criteria for choosing information types and sources
- A process description for each stage of the overall design

Based on these requirements, the overall design of the research conducted is outlined in the following section, before each of the separate sections is described in more detail, and finally validation and ethical issues are discussed.

### 3.4.2 DESIGN EMPLOYED IN THIS RESEARCH

Building on the philosophical perspective and research approach adopted of pragmatism-oriented multi-methods research, this section describes the practical processes and actions taken, and their time frames, towards answering the research questions. As the research was exploratory in nature, the design of the research processes evolved over time, and so this section is a post-rationalisation of the method pursued. Exploratory research aims to go beyond observing and describing phenomena towards uncovering causal links and extrapolations of the impact of this knowledge [545]. Further, as exploratory research, the design progressed through the spectrum of research methods proposed by Demeyer [528], from feasibility study to formal modelling, as far as possible within the limitations of the study, and from an 'exemplar' case study to a proposed benchmark for application layer interoperability.

The overall research design was split into 3 stages, as illustrated in Figure 13. The first stage of the research design was an exploratory literature review to clarify the research questions from a generic research objective. The second stage was then a participatory action research process which engaged with experts through the innovative research, development and testing of semantic web of things systems in 6 projects: 3 in the energy domain, 1 in the water domain, and 2 across domains. Chronologically, these transitioned from investigating the broader impact of semantics in IoT systems through a participatory approach, to investigating the specifics of developing, validating, and leveraging ontologies and semantic web aspects of SWoT systems. The third stage then brought together the findings and analysis of these 6 projects into a unifying design research process, which extended the work conducted beforehand without being participatory in nature. This aimed to bring the separate projects together to provide greater insight into answering the research questions posed, by further developing the artefacts and subsequently testing these as an extension of the state of the art.

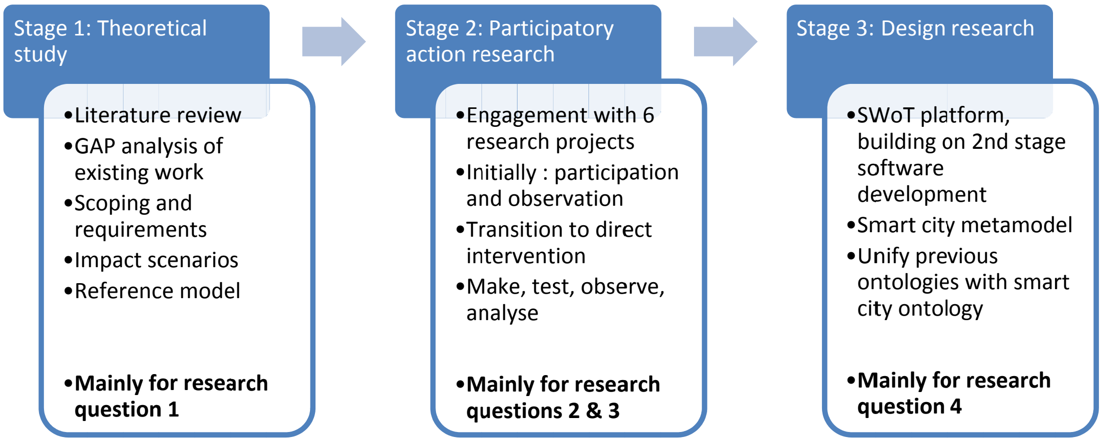


**Figure 14: Overview of the research design**

This literature review began as a very broad consideration of work related to topics in the fields of semantic web, internet of things, and information modelling. As the research landscape and hence problem space became clearer, the research questions were formalised and the required research design was then established. The initial topics considered were IoT, semantic web, BIM, optimisation, cybernetics, artificial intelligence, data management, systems integration, and interoperability. This first stage also involved a thorough evaluation of the existing SDOs, standards, software tools, and best practices available to assist with the remaining research processes. This resulted in a number of impact scenarios and types of ontology and ontology usage patterns, as well as the reference model used to frame the semantic aspects of ICT impact in expert systems.

Following this, the second stage of the research design was to iterate through several relevant participatory research projects, with an increasing shift towards action research at each iteration, and an increasing shift from exploring the broad solution space to specifically analysing the causal links surrounding the research questions posed. This transition from broad to narrow, and from participatory to participatory action research, was in line with the chronology of the research questions, as the earlier research objectives were to understand the broad role of ontologies and ICT, and the requirements they must fulfil, and the later questions aimed to observe the effect of filling these requirements through SWoT, and to investigate causality around the nature of the value presented by ontologies and the semantic web of things. This stage built on the impact scenarios and lessons learnt previously tested by engaging, contributing, and observing in the 6 projects which applied semantic web, IoT, and AI to the built environment. This resulted in a number of artefacts, processes, and further lessons learnt.

Finally, to fully answer the 3rd and 4th, deeper research questions, the latter participatory projects and the third research stage brought together previous work across the separate studies. The third stage extended this within a design research methodology, which aimed to produce a novel solution to established challenges based on the learning to that point, to provide further evidence and deeper understanding around the research questions. Design science utilises the pragmatic problem solving process of design, to serve a research purpose. Kotzé et al. [546] state that the distinguishing feature of design research beyond general design is relevance and rigour. Within the proposed methodology, this involved developing a unifying ontology for smart cities, with a focus on supporting the utility domains researched previously. It also involved developing a semantic middleware platform suitable for utilising the developed ontology in smart city IoT applications. These artefacts were then tested for suitability and extensibility to meet typical needs in terms of big data, privacy and security, and reliability.

As the research contained participatory processes, and a multi-stage approach, it is important to establish the specific contributions to the participatory projects engaged with, and how these contributions led to the overall knowledge outputs. It is also important to reinforce the role of each process within the overall design, in a cohesive, question-oriented manner, as this follows from the philosophical stance adopted of pragmatism. These aspects are clarified in Table 7, Table 8, and Table 9, for each of the 3 stages respectively. Each of the stages is now discussed in more detail, and further details are presented in their respective sections.

**Table 7: Breakdown of research stage 1: theoretical study**

| **Work conducted** | **Role in answering research questions (main role in bold)** |
|---|---|
| -Extensive literature review<br>-Comparative and GAP analysis of existing work<br>-Impact scenario development | **Q1:**<br>-Provided theoretical grounding and rigour to the following work and valuable scope consideration for bounding the decision space<br>Q2:<br>-Provided examples from literature of relevant work in |

| | |
|---|---|
| -Scoping step change which SWoT can achieve<br>-Reference model development as a lens for the further work | the surrounding space, and highlighted the gaps<br>Q3:<br>-The GAP analysis highlighted the potential value to developers and decision makers |

**Table 8: Breakdown of research stage 2: participatory action research project engagement**

| **Project** | **Work conducted** | **Role in answering research questions (main role in bold)** |
|---|---|---|
| KnoHolEM | -Impact scenario development<br>-Technical coordination aspects<br>-Thermal simulation model contribution<br>-Data production, processing, and analysis<br>-Optimised rule development contribution<br>-Guiding and testing AI aspects<br>-Analysing component integration and semantics | Q1:<br>-Scenarios and scoping work provided evidence base and reference point.<br>-Direct engagement with AI and application components informed knowledge management requirements.<br>**Q2:**<br>-Example of a SWoT system architecture in the building energy domain.<br>Q3:<br>-Analysis of the system contributed to evidence base about the role of various components of a SWoT approach. |
| MAS2TERING | -Development of OWL domain ontology based on existing standards and system requirements.<br>-Contribution to metaprogramming to convert OWL ontology to JavaBeans ontology, for | Q1:<br>-Some further evidence of requirements on semantic components in SWoT systems for smart grids, broadening evidence base.<br>**Q2:**<br>-Example of a SWoT system |

| | | |
|---|---|---|
| | deployment in MAS.<br>-Contribution to deployment of ontology in MAS.<br>-Analysis of resultant system | architecture in the smart grid domain. Different use of explicit semantics provided breadth to evidence base.<br>-Direct contribution to ontological aspects provided depth to the evidence base.<br>Q3:<br>-Analysis of resultant system and experience in development provided breadth to the evidence base. |
| RESILIENT | -Contribution to OWL domain ontology<br>-Consultation with domain experts about role of the SWoT system<br>-Contribution to deployment of ICT system<br>-Analysis of resultant system | Q1:<br>-Some further evidence of requirements on semantic components in SWoT systems for polygeneration grids, broadening evidence base.<br>**Q2:**<br>-Example of a SWoT system architecture in the polygeneration grid domain. Again, different use of explicit semantics provided breadth to evidence base.<br>Q3:<br>-Analysis of resultant system and experience in development provided breadth to the evidence base. |
| WISDOM | -Primary contributor to development of scenarios, use cases, sequence diagrams, system requirements and decomposition to component requirements.<br>-Consultation with domain | Q1:<br>-Evidence on SWoT system requirements in water domain, adding significant breadth<br>Q2:<br>-Extensive depth added to evidence base from leading comprehensive semantic modelling task |

| | | |
|---|---|---|
| | and ICT experts about role and value of ICT and SWoT in target systems.<br>-Sole contributor to ontology scoping, requirements, development, validation, testing, and deployment<br>-Sole contributor to semantic web software and semantic inference software development<br>-Integrated developed software into wider smart water ICT platform | -Breadth added from different mechanisms of using explicit semantics to other projects.<br>**Q3:**<br>-More extensive expert consultation and testing of SWoT approach provided significant evidence towards this question<br>Q4:<br>-Some evidence of generalisation across domains, as water domain has significant differences to the energy domain |
| CUSP | CUSP-energy:<br>-Contributed to semantic web software, ontology deployment, and guided the leveraging of the SWoT approach<br>-Contributed to system architecture design, and GUI development<br><br>CUSP-water:<br>-Developed a GUI which tested and ultimately demonstrated the value of the approach<br><br>-Analysis of system development and resultant | Q2:<br>-Example of SWoT system with ambitions across domains, adding further breadth to the evidence base<br>**Q3:**<br>-Adopting role of application developer offered further insight into the value of a SWoT approach<br>-Analysis of system contributed breadth to evidence base<br>Q4:<br>-CUSP generalised across energy and water domains, and had ambitions to be extensible into other domains |

| | | |
|---|---|---|
| | artefacts | |
| Hypercat | -Engaged and consulted with experts from across 12 'verticals'<br>-Closely engaged and consulted with water 'vertical' partner to complement learnings from WISDOM engagement<br>-Extensive literature review and comparative analysis of IoT standards landscape from technological and policy perspectives | Q1:<br>-Literature review, comparative analysis, and expert consultation provided breadth and depth to evidence base<br>Q2:<br>-Provided examples of more lightweight approach to application layer interoperability, where semantics are not handled verbosely<br>Q3:<br>-'Pure' IoT approach demonstrated some benefits and limitations of this more lightweight option<br>-Expert consultation provided significant breadth to evidence base on the role and value of SWoT and IoT<br>**Q4:**<br>-Analysis highlighted the nature of extensibility and interoperability across domains.<br>-Project acted over 12 smart city 'verticals' with over 1100 individuals, so added significant breadth to the evidence base for this question<br>-Emphasis on accessibility contributed to consideration of nature of supporting further work |

**Table 9: Breakdown of research stage 3: unification and generalisation through design research**

| **Work conducted** | **Role in answering research questions (main role in bold)** |
|---|---|
| -Building and testing SWoT platform reference implementation<br>-Curating metamodel for smart city domain<br>-Aligning previous ontologies with smart city ontology | Q2:<br>-Further depth and breadth to evidence base by building on previous stages to produce a modular solution which meets the collective requirements<br>Q3:<br>-Further exploration of the potential value by building on the positives and mitigating the negative aspects of solutions from previous stage<br>**Q4:**<br>-Significant breadth and depth added to evidence base by pursuing the unification and generalisation of the learnings and artefacts from the previous stages to support further work |

### 3.4.3 STAGE 1: THEORETICAL STUDY

The theoretical study aimed primarily to collect and analyse the majority of the evidence base for the first research question, whilst supporting the evidence bases of the other questions. It also served to clarify the research questions, and determined the emerging landscape and resources evident for furthering the remaining research. A significant portion of the evidence towards answering the 1st research question emerged from the literature review conducted during this stage. A key initial output of the theoretical study was an early version of the proposed reference model. This reference model was then used as a lens for the identified literature, within the existing technological landscape in each smart city domain, to identify potential scenarios where semantics may have impact. Firstly, this identified the challenges faced and how ICT and existing ICT trends could help overcome these. This highlighted the areas where relying on ICT required comprehensive semantic interoperability; at the instances of most utilisation of machine traversal of the latter stages of the reference model in a domain.

The identified areas formed the basis of impact scenarios; descriptions of instances where semantically-enabled advanced applications could have impact in a smart city or industrial system. Finally, these scenarios were explored more to identify how semantics could support ICT impact in various usage patterns. This formed a foundation for the experimentation and exploration in target domains in the second stage of the research. The domains of energy and water were chosen as two of the most critical, and for their similarity both technologically from an abstract level, and in a business sense given the utility-centric nature of the industries. The model could equally be explored in other domains, and this is discussed fully in later sections. The exploratory process of the literature review was described more fully in Section 2 alongside the review itself, and the process and outputs of the theoretical stage are discussed further in Section 4.1.

### 3.4.4 STAGE 2A: PARTICIPATORY RESEARCH IN ENERGY DOMAIN

The second stage of the research design was a process of iterative learning cycles, each consisting of a participatory research approach, engaging with a collaborative research project. Each of these was essentially an instance of Kolb's experiential learning cycle [335] whereby one senses the domain, actors, and goals, then tries to understand more about what is required and how others have done that, before deciding on possible approaches. These decisions are then acted on, before analysing the appropriateness of the actions and continuing to iterate to produce learnings whilst also addressing the immediate problem, a key aspect of the action research approach adopted, as discussed previously.

Initially this took place within the energy domain, and then in the water domain. The energy domain was ideal, as a domain where semantic modelling and the value of IoT is already emerging in an ad hoc manner, as this provided a grounding to answering the 2nd and 3rd research questions before then testing these findings in another domain. The resulting artefacts and lessons were extended and tested further in the water domain, through direct development and testing of software and knowledge modelling artefacts, using the project's data and also knowledge gained from consultation with domain experts.

Within the energy domain, the approach was to first consider the manifestation in a relatively less complex case; at the building level. This allowed greater isolation of the effect of utilising ontological representations and semantic web technologies. This was then expanded to consider groups of buildings within a smart prosumer grid. This tested the hypothesis across further experimentation. Finally, greater complexity was then added by considering additional energy vectors at the multi-building scale. Specifically, this involved the utilisation of semantic web and IoT technologies within an energy hub decision support tool, which involved the optimisation of power and district heating systems in parallel at the supply side, based on simulation and a wide range of heterogeneous data sources and remote resources. Whilst the role of semantics is already emerging in the energy sector, it is still embryonic in the water sector. Therefore, the water sector was chosen for an in-depth case study, to observe the replicability of the benefits observed in the energy sector in other types of systems, with water systems chosen due to their partial similarities to energy systems. This is discussed further in Section 3.4.5, after each of the participatory energy research methodologies is discussed in more depth.

#### 3.4.4.1 SMART BUILDINGS

Within the second stage of the overall research design, the hypothesis and reference model was first explored at the building level, with a focus on energy management. This was achieved by engaging in and contributing to an ICT solution which achieved semantically-enabled advanced application decision support. The aim of this process was to elicit significant lessons about exploiting semantics in supporting the ICT traversal of the latter stages of the reference model. This began to prepare an evidence base for the second research question, and also added real world experience to the theoretical evidence towards the first research question. This research was conducted within the context of the EC FP7 research project KnoHolEM. The overall approach of the project was to integrate rules derived from empirical and theoretical analyses with a user interface and real world sensors and actuators through a semantic web approach, as shown in Figure 14. The main engagement with the currently described study was within the theoretical analysis shown in Figure 14, by using the project to assist with scenario definitions and energy modelling, and then contributing to the other theoretical analysis activities and integration with the further systems, with more details shown in Figure 15.

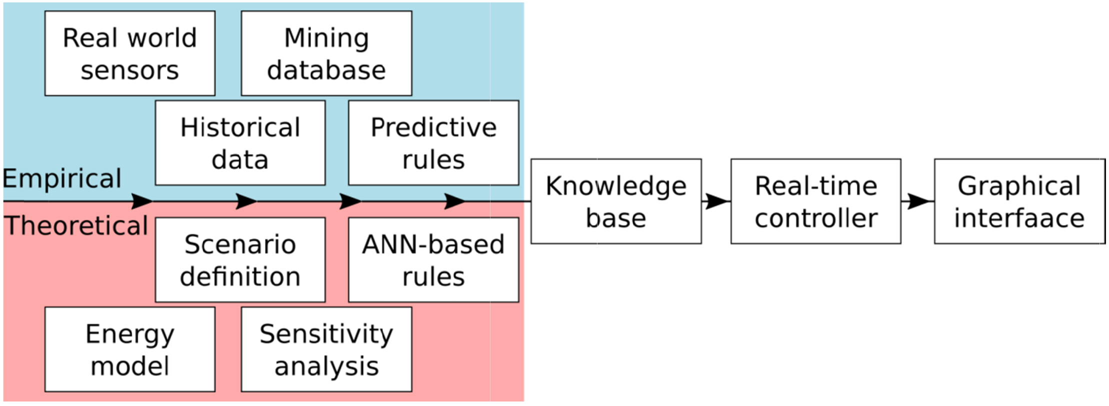


**Figure 15: Overview of the KnoHolEM BEMS approach**

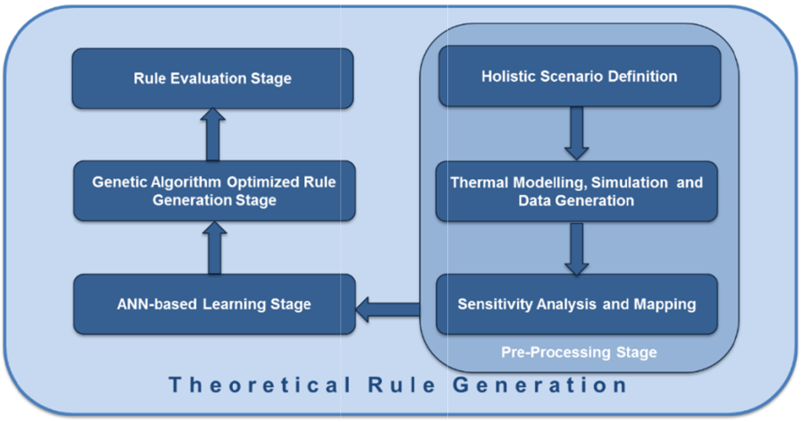


**Figure 16: Processes in the generation of theoretical rules**

The KnoHolEM project involved the installation of a relatively small number of wireless sensors into pilot buildings, to detect properties such as temperature, air speed, occupancy, and humidity, as well as web-enabling existing BEMS systems and sensors and actuators, and BIM models of the buildings. These were then communicated to a cloud-based retrofit building energy management system developed within the project. The solution was primarily tested within a mixed mode residential care home in the Netherlands. The primary goal of the work was to produce a BEMS, which could be retrofitted into public buildings with minimal investment to ultimately improve energy consumption with a lower capital expenditure, hence improving the ROI of the solution. This involved a semantic knowledge base, and novel analytics which included the automated production of rules, fuzzification, and data mining on historical metering data. The visualization component delivered engaging 3D WebGL and energy performance monitoring and decision support. The BEMS aimed to promote trust with facility managers (FM) through a negotiation based user-in-the-loop approach.

The current work engaged with and observed the described project as a case study of the application of advanced applications and semantic technologies in the built environment, specific to the domain of building energy management. This was then compared to experience in the other case studies towards the development of the generic framework proposed. Especially, this was compared to the aspects of the other case studies which were relevant at the building level; distributed generation and load flexibility optimisation in case study 2, and domestic water behavioural change in case study 4.

The case study involved active engagement with the described project. This primarily involved the development of the impact and optimisation scenarios through systems analysis and expert engagement, as well as collaboratively developing the simulation models and implementing the simulation-based rule generation process in the described building. This exposure to various activities in the research, development, and testing of semantically-enabled advanced applications allowed an analysis to be conducted of the project for the purpose of the current thesis, from which the lessons learnt are drawn. As this stage of the research occurred chronologically alongside and shortly after the literature review, this stage formed further preliminary knowledge towards answering the first 2 research questions. Whilst the research process did not involve directly contributing to the semantic web components, it leveraged the literature review lessons directly to produce impact scenarios, which forms part of the research contribution proposed, and to gain an understanding of the artefacts and processes utilised in SWoT in a far more valuable manner than pure literature review, as it allowed monitoring and evaluation of the process and impact of leveraging ontologies and SWoT technologies in real world systems.

#### 3.4.4.2 SMART PROSUMER GRIDS

Experience from the previous building energy project was used to explore the hypothesis and reference model in a more complex system: the management of smart prosumer grids. A semantic model was developed to support the ICT traversal of the latter stages of the reference model through an agent-based approach, coupled with weather-based predictive capabilities. The lessons learnt previously were used to develop the model, then its usage patterns were observed, as well as its role in the ICT solution and impact on the industrial system. Again, lessons were learnt and outcomes were formalised. This primarily aimed to provide evidence towards the 2nd research question, as it involved developing ontologies and integrating a semantic web approach with a MAS in a smart home and smart grid system. The analysis of this work then contributed to the 3rd research question. The project consortium developed a MAS and web-service system for optimising the load and energy storage scheduling of smart grids, based on the arrangement presented in Figure 16. Integrating these agents with each other and the web services developed required a comprehensive approach to knowledge management and interoperability, which was the aspect engaged with for the purposes of the currently described study.

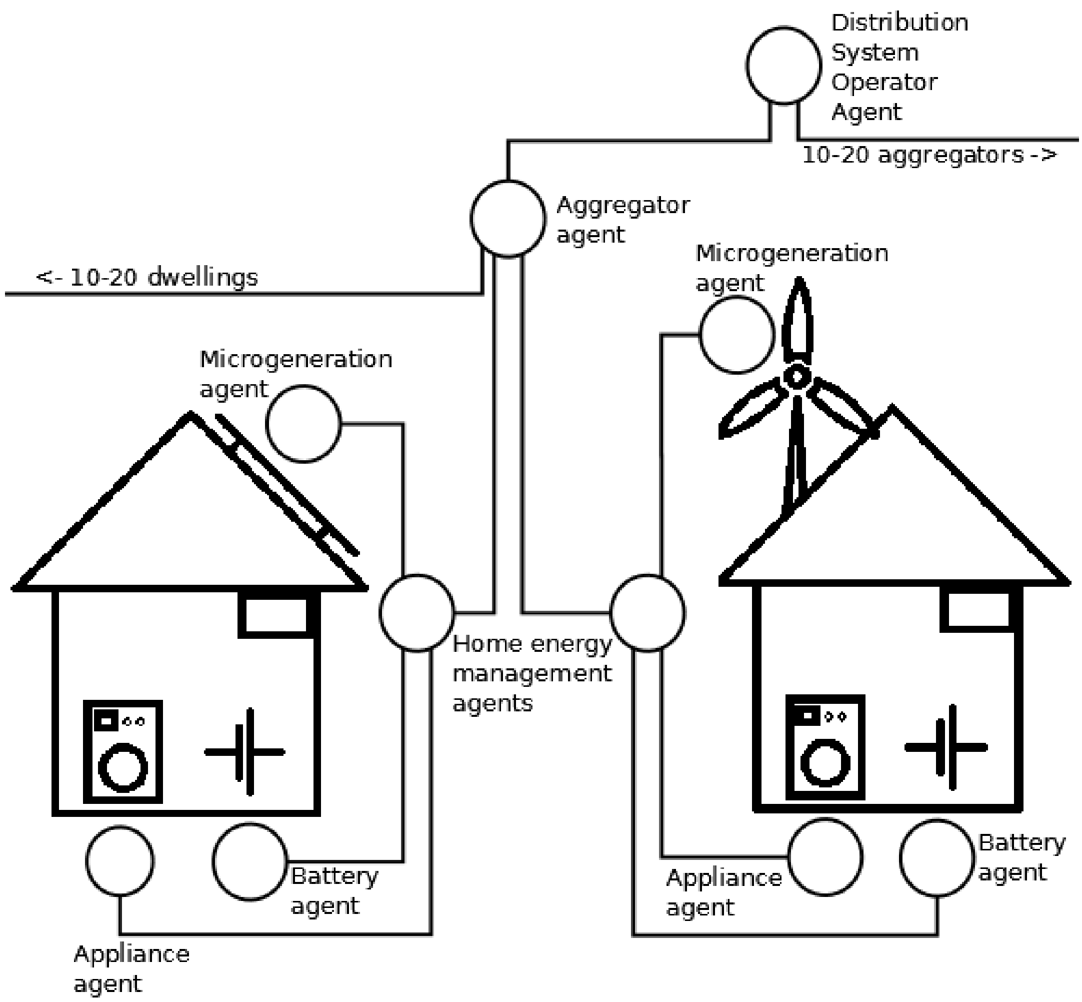


**Figure 17: Illustration of smart home-smart grid MAS**

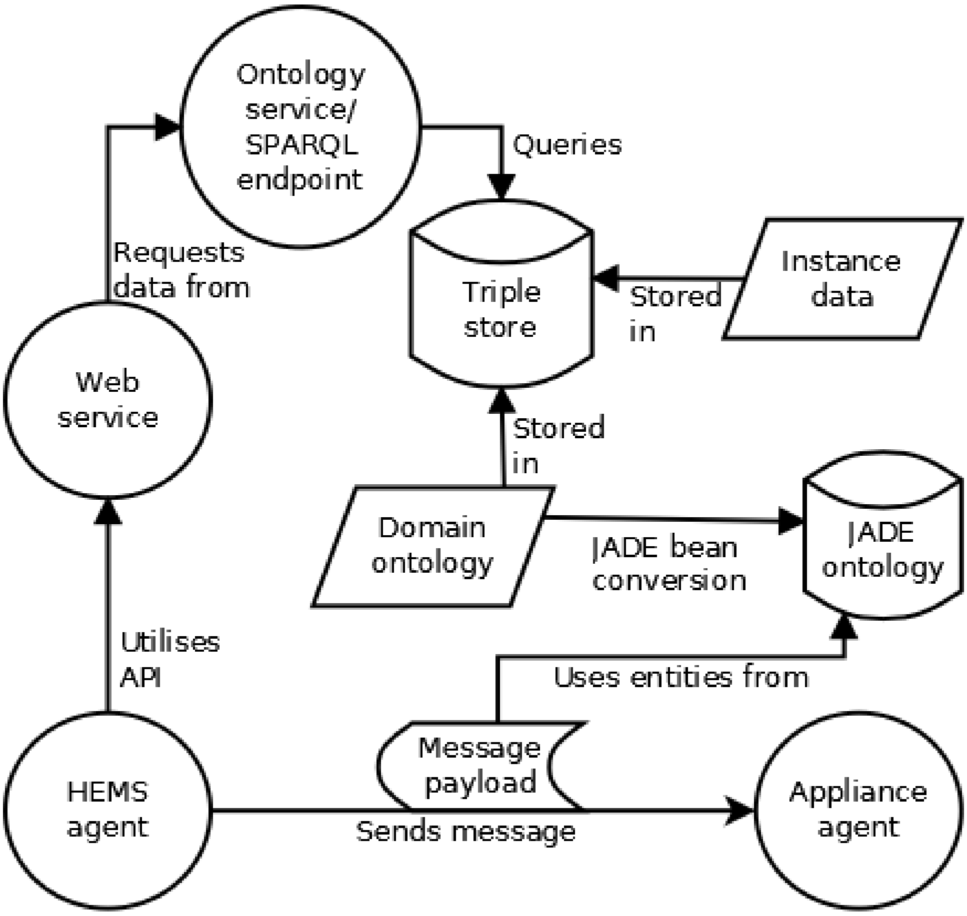


**Figure 18: Knowledge management approach developed for smart home-smart grid system**

The previous work on energy management at the building level was extended to their wider power network, in the emerging energy context of prosumers and microgrids. This case study focused on the work conducted and analysis of the EC FP7 collaborative research project entitled 'Multi-Agent Systems and Secured coupling of Telecom and Energy gRIds for Next Generation smart grid services' (MAS2TERING). The project aimed to develop a MAS which was capable of optimising both domestic demand and energy generation and storage vectors through the emergent trading of the virtual commodity of flexibility, thereby creating and testing new markets and business models which supporting efficiency and grid resilience. The described MAS was also coupled with novel prosumer forecasting services, and the entire solution utilised a shared ontology and data model to facilitate communication.

The current work utilised the described project as a participative action research process, through which various aspects of the hypothesis were tested in a different manner to that of the previous case study. Specifically, this case study involved the development of an ontological solution to enable the traversal of the reference model's stages through ICT, and collaborative deployment of the solution within the ICT system, followed by analysis and consideration of the process, and artefacts developed.

The development of the ontology firstly required a thorough scoping and knowledge gathering process to bound the domain to be modelled. Lessons from the previous case study were also used to guide the process, to produce more suitable requirements for the ontology, and to use a more collaborative and balanced approach to the development process. After conducting the knowledge gathering and scoping process, the ontology curation was conducted manually. After reusing existing semantic resources, concepts specific to the target system were then included in a coherent manner. Further relationships, data properties, rules, and restrictions were then added, to provide a comprehensive domain perspective and data model. The ontology was then integrated with the knowledge management system developed by experts within the project. This involved collaboratively developing a meta-programming tool to convert the ontology to a JADE-compliant JavaBeans ontology, and by building the accompanying software infrastructure to host and query the artefacts.

#### 3.4.4.3 SMART POLYGENERATION GRIDS

The inclusion of heat as an energy vector in districts is growing in popularity, and adds significant complexity to the system of systems. This was therefore chosen as the next case study, where once again, semantically-driven advanced applications provided decision support to assist the traversal of the reference model. This aimed to build on the previous case study to provide further evidence towards the 3rd and 4th research questions, by contributing to and evaluating the performance of ontologies and SWoT systems in the domain. Lessons from the previous case studies were used to contribute to the semantic aspects of the ICT solution, and again the role and impact of the semantics were observed. Each of the case studies adopted different usage patterns for semantics and IoT, in different applications, and so significant breadth of knowledge was achieved. Further, as this case study held more complexity in the application domain, due to the multiple energy vectors, it also provided more depth of evidence, by testing the hypothesis and reference model in more challenging ways.

The case study focused on work conducted in the context of the EC FP7 project entitled RESILIENT. This project aimed primarily to optimise the generation mix within a district polygeneration grid. Specifically, the project’s intervention acted at the district’s energy hub, which included both power, heat, and cogeneration energy producing units. The energy hub served power to nearby public buildings through a low voltage grid, and was connected to the national grid. The energy hub also stored heat energy, and distributed heat to the nearby public buildings through a district heating network. The project therefore oversaw the installation of new renewable energy and storage units, and aimed to optimise their positive impact on the performance of the system. This was achieved by thermal modelling the public buildings, which allowed predicted demand profiles to be produced. These were then utilised within a multi-objective optimisation algorithm to produce a pareto-front of setpoint options for the district managers to choose between. The solution utilised a semantic web approach to integrate the various components, including the storing and visualising of BIM models, the instantiation of simulation and optimisation models, and the discovery of sensors.

As with the usage of the MAS2TERING project, the current work utilised the RESILIENT project as a case study, through which various aspects of the hypothesis were tested in a different manner to that of the previous case studies. This case study involved contributing to the development of the ontological solution described, and collaborative deployment of the solution within the overall system described, followed by an analysis of the evidence collected. The ontology scoping and development used the methodology shown in Figure 18

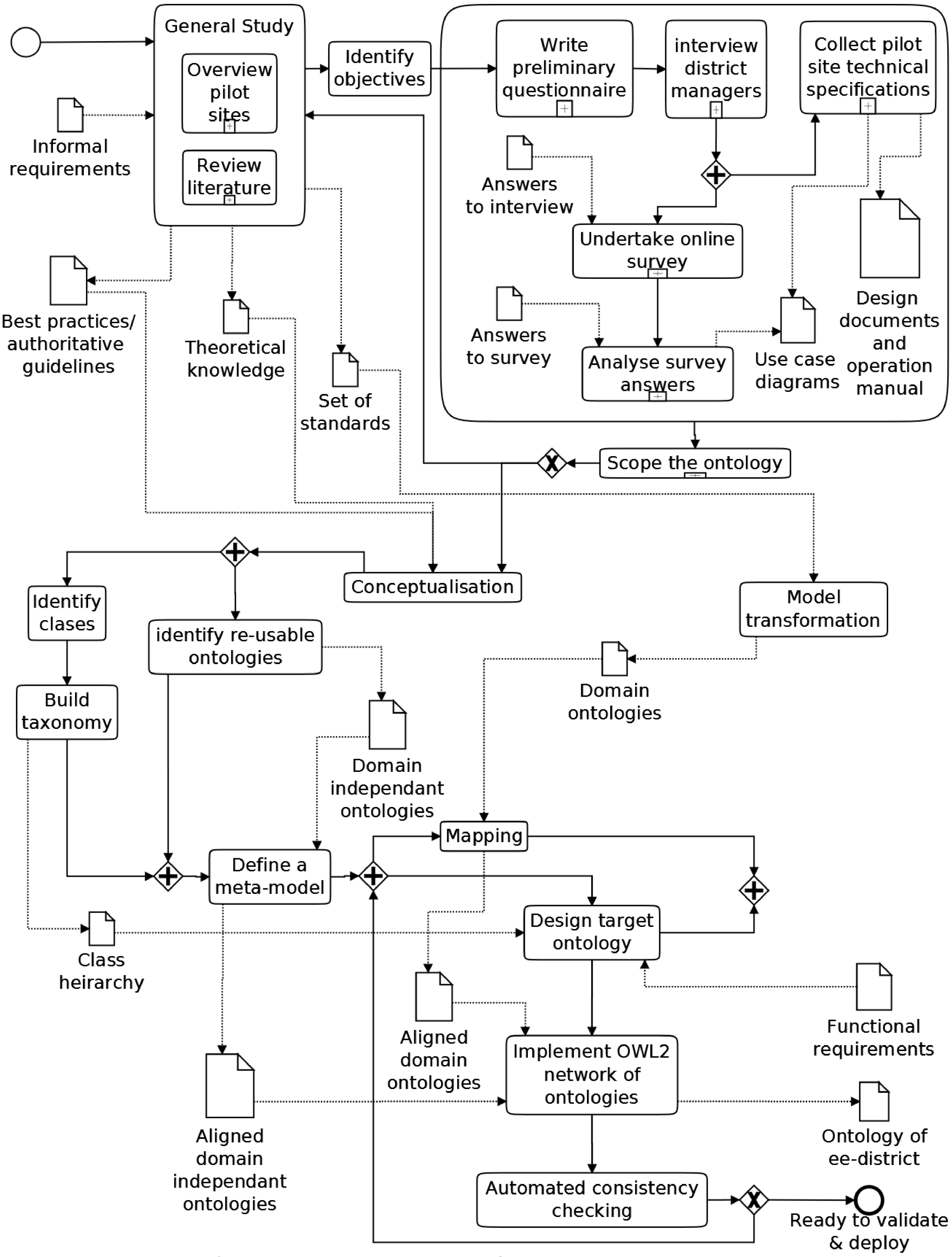


**Figure 19: Methodology for ontology development in RESILIENT [547]**

## 3.4.5 STAGE 2B: PARTICIPATORY RESEARCH IN WATER DOMAIN

The knowledge acquired from the previous case studies sufficiently justified the role of semantics in energy-centric systems. This section describes the work conducted to extend this evidence to the water sector, and with more control over architectural and software decisions and development. An in-depth case study was pursued through the scoping, developing, validating, testing, and exploitation of semantically-enabled services for the water sector. The work was conducted in the context of an EC FP7 project, entitled 'Water analytics and Intelligent Sensing for Demand Optimised Management' (WISDOM).

The WISDOM project aimed to integrate legacy sensors, existing data repositories, and new sensors, through a cloud-based service oriented architecture, and to demonstrate the value of this through several intelligent applications. This included a hardware and communications layer, a core services layer, and an applications layer. The core services layer served as the primary component of interest to the case study, although its interactions with the application layer were also relevant. The core services layer included an event-based knowledge management solution alongside analytics, rule-based, and optimisation services. These were integrated through the use of a semantic web approach, grounded in the domain ontology developed within this case study. The project involved several industrial partners, and tested the proposed solution within the business and technological contexts of those companies, in 5 pilot sites across Wales, France, and Italy.

Similar to the previous two described energy case studies, the current work utilised this project as a case study, but through more direct engagement with the semantic aspects of the work. Specifically, the entire ontology and semantic web software scoping, development, and testing, was performed within the context of the current study as the sole contributor. This allowed more detailed control over the processes undertaken and a more thorough experience of both the process and the results. This provided significant exposure across the lifecycle of semantic artefacts in their role within an industrial smart city system. As semantic modelling in the water industry is embryonic, lessons learnt through this case study were deemed more likely to be applicable to other domains, where semantics are equally novel.

The methodology adopted in scoping, developing and deploying the WISDOM semantic models primarily utilized the recommendations of the NeOn methodology to utilize a collaborative approach with domain experts and ontological experts, through an iterative process shown in Figure 19. The main stages involved in this workflow are described through the following sub-sections. Firstly the knowledge acquisition and scoping phase is discussed, then the development of a domain independent meta-model as an extension of reusable ontologies, then each of the stages of the actual ontology development process are elaborated. The pilot site instantiation process is then discussed before the web service development process, the preliminary validation, inference engine development process and secondary validation and finally the process of mapping to other ontologies.

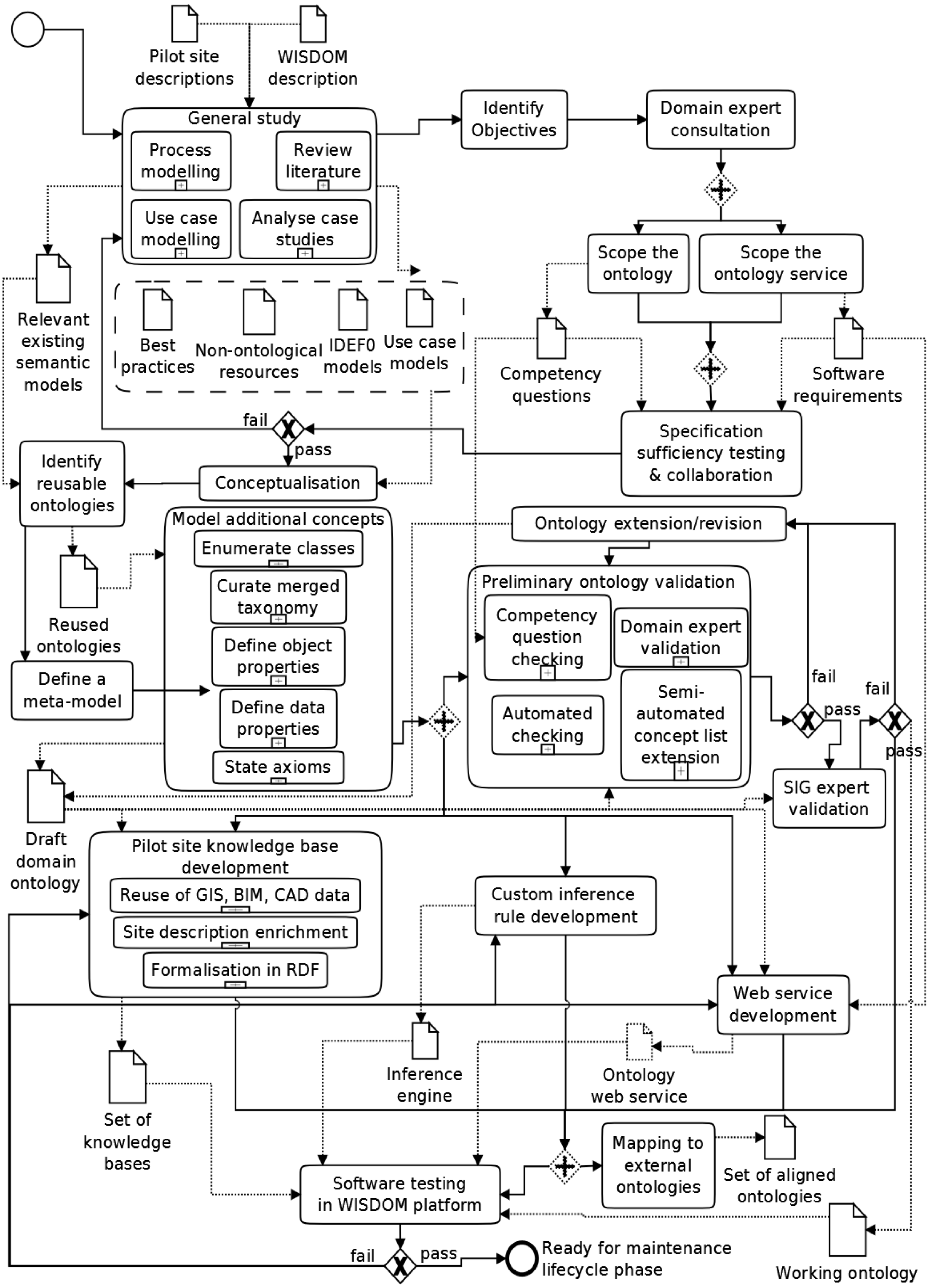


**Figure 20: WISDOM ontology and knowledge-management system development processes**

#### 3.4.5.1 REQUIREMENTS ENGINEERING

The first phase of this learning iteration was to thoroughly understand and bound the challenge to be faced in the clients' water utility management systems, through a comprehensive requirements engineering process. This was loosely comprised of 3 stages, as shown in Figure 20. After gaining a conceptual understanding of the domain, pilot sites, and business processes, through literature review and expert consult, these were formalized into IDEF0 [548] models, use case models, explicit scenarios and deployments, and finally, software requirement specifications. These were all iterated through a collaborative process with domain experts to promote their accuracy and completeness.


**Figure 21: High level stages of the smart water requirements engineering**

The requirements engineering process modified NeOn to account for the growth of the Internet of Things, and to promote reuse in the field. This involved balancing the NeOn's knowledge engineering objectives with software engineering objectives, and the softer requirements of fostering client 'ownership' and human intelligibility. Also, the knowledge gathering stage was supplemented with a semi-automated web-based process of concept extraction, as described in section 3.4.5.1.4, which primarily assisted in developing the semantic water models.

After knowledge gathering and scenario specification, an analysis and design process was followed to produce software requirements for the overall software solution. These were then iterated alongside domain experts, and a system architecture was curated, before the requirements were decomposed for each component. The software requirements were then decomposed further to produce a set of ontology competency questions.

Once the requirements had been developed, they were iterated against meta-requirements. This promoted coherency across the three perspectives of i) software specification, ii) knowledge modelling, and iii) domain reuse, after which the development of the ontology and accompanying software was undertaken. The requirements were used throughout the process to test and guide the software and ontology developments. The following sub-sections describe the requirements engineering steps in more detail.

##### 3.4.5.1.1 SCENARIO IDENTIFICATION

The first milestone of the requirements engineering process was to produce platform-level impact scenarios. These described the various impact pathways for the software within the existing business processes and software frameworks present in industry. The methodology consisted of four stages, as shown in Figure 21, conducted in close collaboration with the industrial stakeholders. This aimed to foster early practitioner engagement with the developed artefacts, 'buy-in' of domain experts, and genuine business and industry value.

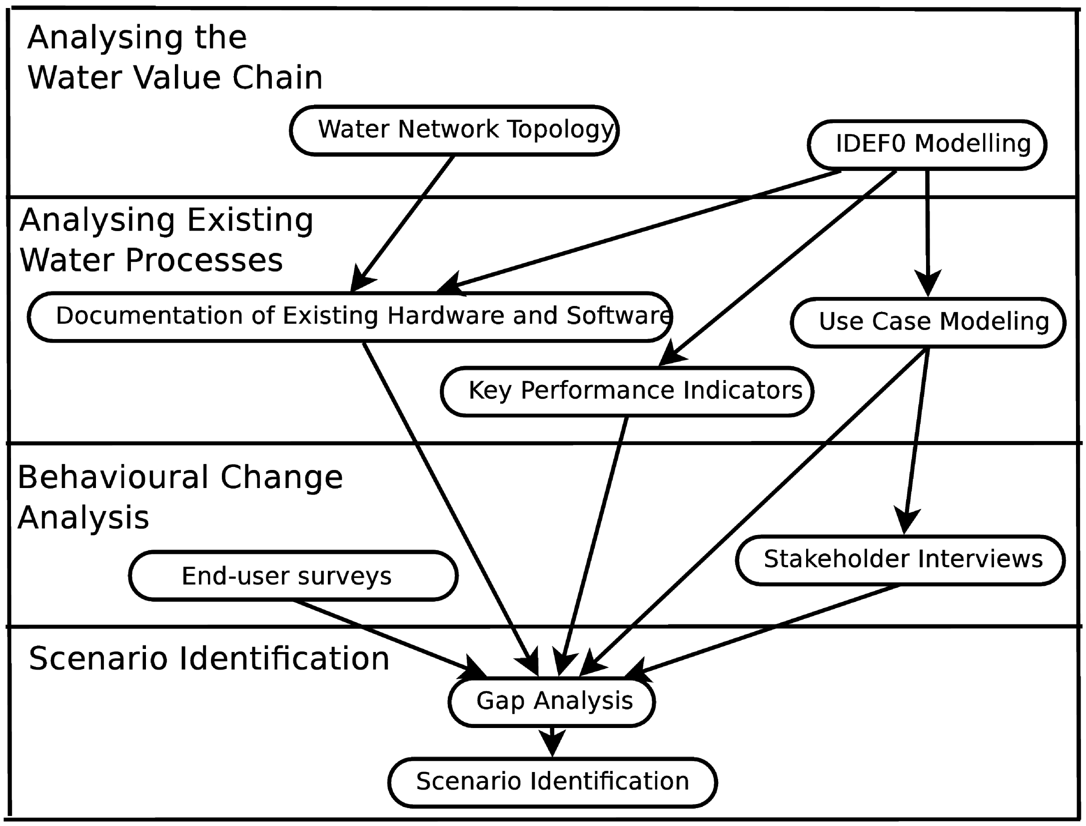


**Figure 22: The WISDOM Scenario Identification Process**

The development of impact scenarios began with pilot site system specification and business process modelling. Next, informal knowledge gathering was conducted through expert consultation, literature review, site visits, and analysis of the existing products and processes at the client organisations. Use case modelling of the existing system, and KPI documentation was conducted alongside this. Finally, this knowledge was unified into a gap analysis which resulted in a number of scenarios being identified. For each scenario the following fields were populated:

- Name
- Description
- Objectives
- Artefacts to be developed
- Input Data
- Existing Technologies to Utilise
- Output Data
- Actors (during demonstration and at other times)
- When Applicable
- Anticipated Impact

Once generated, scenarios were reviewed and iterated until a final set of scenarios, sufficiently covering all targeted aspects of the water value chain, was identified. The main tasks in this stage of the requirements engineering process are described in the following subsections.

#### 3.4.5.1.2 SMART WATER DOMAIN ANALYSIS

The first stage of the requirements capture process involved achieving a high level understanding of the structure and the processes involved in the water value chain from industrial experts. To achieve this, the first stage is broken down into two tasks: a) Documenting water processes using the IDEF0 [548] functional modelling methodology and b) the analysis of network topology specifications.

In order to produce IDEF0 models for each pilot, the system within each pilot location was analysed and the following tasks performed:

1. Document the high level processes that the water goes through within the pilot.
2. For each process identified the inputs and outputs must be identified and using these inputs and outputs processes are connected.

3. For each process identified the constraints (standards/legal requirements, economic frameworks, quality/quantity requirements) and mechanisms (actors, existing software, existing hardware) must be identified.
4. Once the high level model of the system has been produced, each process on this model should be broken down and the IDEF0 modelling process is repeated for each sub-process.

##### 3.4.5.1.3 ANALYSIS OF CLIENT'S SYSTEMS AND PROCESSES

The second stage of the requirements capture methodology builds on the understanding of the pilots' water processes and topology. This stage consists of three tasks: a) Documentation of existing hardware and software used within the pilot b) Documentation of key performance indicators and c) UML (Universal Modelling Language) Use Case Modelling [549].

The first task involved identifying further information for each of the mechanisms identified as part of the IDEF0 modelling. To achieve this, a template was provided to the client, asking for information such as name, type, data storage technology, and file format.

The second task involved specifying in more detail the key performance indicators for the various processes within the pilot, the majority of which had been identified as constraints during the previous stage of the process. The final task in the second stage was to understand the interactions of actors with the water system. To achieve this, a series of use case modelling exercises were conducted. The IDEF0 models were analysed and all actors that featured as mechanisms were used as a starting point for generating use cases. These described in a standard notation the interactions of individuals with the target system.

##### 3.4.5.1.4 SEMI-AUTOMATED WEB CRAWL AND FEATURE EXTRACTION

As supplementary work in gathering knowledge about the target domain, the manual elicitation of domain knowledge was coupled with a semi-automated web crawl and feature extraction process. The aim of this was to facilitate broader relevance of the ontology by aligning the terminology and semantics modelled with the wider water sector, by analysing web documents across the sector and ontological features from these, as a whole body of literature. The output of this process was a list of the main smart water concepts used in web documents on the web sites of relevant organisations such as utility companies, education bodies and regulators.

As a supplementary process, the full details of the semi-automated validation are outside the scope of this thesis, but it is briefly summarised here. The first stage consisted of automatically 'crawling' a manually selected list of relevant websites (and their linked websites) for public HTML data (raw 'screen text'), Microsoft Word, .txt and .PDF documents based on loose rules for relevance, through a custom-made Python program. These were then processed to extract a list of all the words, and several metrics about them, per document. This data was then further processed through another novel Python program to identify the most relevant and likely candidate class and property names.

The pertinent term extraction program developed first ranked the words by frequency and 'term frequency, inverse document frequency' (tf-idf), a common metric of the importance of a word in a document. The program then filtered out 'stop words' such as 'if', 'the' and 'is', then looked up the word in the WordNet lexical database and retrieved a definition of the word. The relevant words in this definition were then looked for in the list of words found in the crawled web documents, and their tf-idf values summed when they were found, to produce another metric of 'importance' for the crawled document word in the target domain. This was then utilised alongside each word's tf-idf value to produce a hybrid measure of importance of the word, and the master list of terms was ordered by this metric of importance.

Finally the script utilised WordNet to separate the proper nouns (which would be instances of classes), nouns (which represent candidate classes) and adjectives and adverbs (candidate properties); other linguistic components were removed. From the resulting data an indicator of the domain coverage of the ontology was calculated when validating the ontology and possible missing classes and properties were identified. The output of this process complemented the manual requirements engineering stage, primarily by promoting better domain coverage and relevance of the ontology towards reusability.

### 3.4.5.1.5 SOFTWARE REQUIREMENTS ELICITATION

The described impact scenarios served as a guiding set of initial intentions, which represented a project decomposition and description task, and somewhat represented client requirements. From these, a set of software requirements were curated through system analysis and design processes, with domain expert consultation, as shown in Figure 22. This process was conducted using a hybrid top-down and bottom-up approach, because whilst the majority of the scope boundaries were established in the scenario descriptions, end-user and data-oriented perspectives were also important. Water consumer views, external expert validation, and descriptions of the available data and hardware, were also used to define the required system.

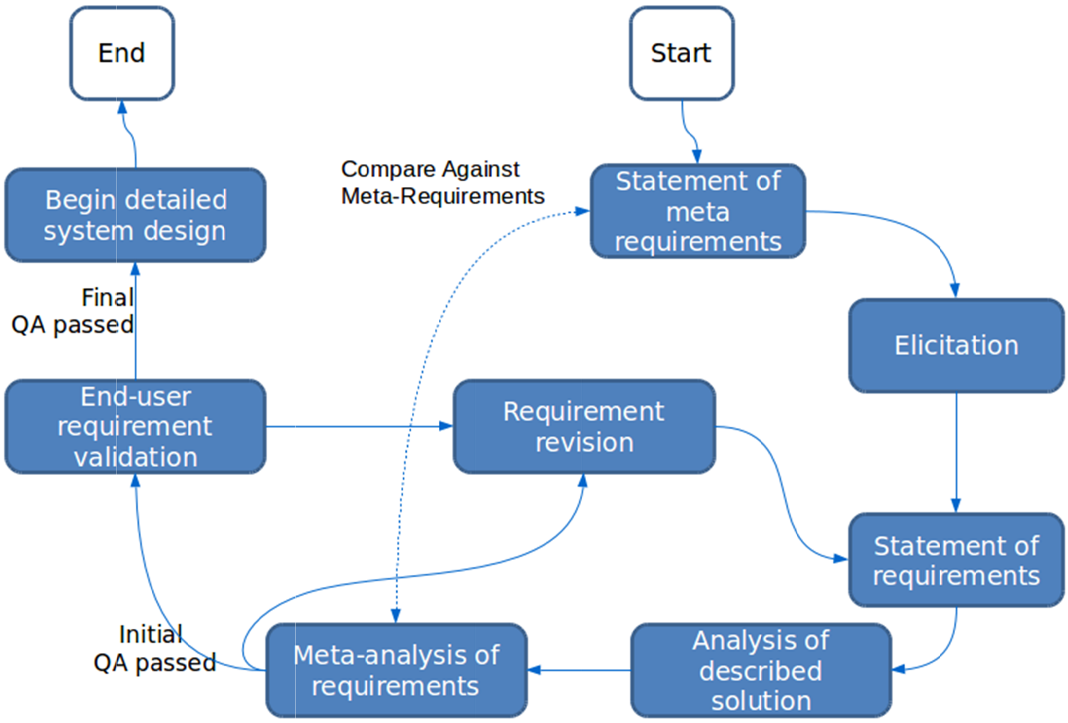


**Figure 23: Requirement engineering processes following scenario identification**

The first step in this process was describing a set of meta-requirements based on best practice and expert knowledge amongst the partners. These meta-requirements defined the guiding principles by which the quality of the solution's requirements can be verified. The elicitation process then involved an iterative brainstorming process. At each stage, scenario use case diagrams were produced for the envisaged software. These diagrams were then formally compared with each scenario's goals and subsequently revised. This represented a dependency model and provided accountability for the requirements engineering. The client was iteratively consulted as part of the requirement elicitation, as described in the next subsection.

As well as comprehensively eliciting functional requirements, a set of non-functional requirements elicitation questions established by Michigan State University [227] was used to assist with gathering additional non-functional requirements. This produced requirements related to the quality of the functions delivered, such as response times and required ease of use.

The end result of the initial requirements elicitation process was a set of definitive statements of requirements. However, these requirements were varied in terms of terminology, depth of specification and compliance with the meta-requirements, as they were the result of an organic and multi-perspective elicitation process.

#### 3.4.5.1.6 REQUIREMENTS ANALYSIS AND REFINEMENT

Following the elicitation and gathering of initial requirement statements, it was necessary to thoroughly analyse and revise them. The initial brainstorming approach produced a scenario-based list of requirements, but the terminologies and levels of detail of these were inconsistent due to the multiple perspectives adopted. The initial requirements were improved and homogenised by abstracting them from their scenario specific contexts and considering the entire set as a description of the overall system. This enabled the initial validation and improvement of the requirements as omissions, duplications, ambiguities and variations in terminologies were exposed.

The main task that facilitated the analysis of requirements was explicitly mapping between the 13 discreet scenarios and the functional requirements. This checked that the requirements meet the scenario and project goals, and also check the relevancy of each functional requirement, to avoid over-specification. As well as dependency mapping, developing sequence diagrams to illustrate how the defined system functions achieved the scenarios highlighted several opportunities for improvement. The requirements were next compared against the meta-requirements to ensure they were of a high enough quality, completeness and testability to sufficiently describe the system requirements. The final stage of validating the requirement specification was consultation with the end users and system designers.

###### 3.4.5.1.7 DECOMPOSITION TO ONTOLOGY SOFTWARE REQUIREMENTS

Producing the requirements of the ontology as a knowledge modelling artefact was a significant part of the requirements engineering. As ontologies aim to progress towards domain consensus, it is preferable that they not only meet system specific objectives, but balance this need with the goal of achieving an agreeable, complete, and sufficient representation of the domain. This is critical in the emerging field of urban cybernetics, as the initially intended system is highly likely to evolve to be integrated with external systems, new system-level functionality, and the ontology itself may be reused elsewhere, so it is beneficial if it is suitable as such. This involved additional requirements, and closely engaging with practitioners from an early stage.

Once the system-level requirements were deemed sufficient, these were decomposed further into component-level requirements, including the knowledge management software service based on the ontology. These software requirements, alongside the scenarios, automated term list, reusable ontologies, and elicited domain knowledge, were used to produce a set of competency questions to bound the scope of the ontology. Again, these were iterated alongside the other requirements engineering tasks and mapped against the scenarios' main entities and pertinent data, to promote completeness. Finally, the competency questions were converted into a set of SPARQL queries to serve as a litmus test, although the set of questions evolved through the project, as the role of the ontology became clearer.

.

###### 3.4.5.1.8 PROMOTING ONTOLOGY REUSE

One ambitious goal of the task was to contribute to the relatively new discourse in the water sector regarding semantic modelling and standardisation. Whilst a somewhat secondary goal, it was deemed worthwhile and feasible given the novelty of the concept in the water sector. Arguably this is also true in other smart city domain, such as smart government, smart food management, and smart mobility.

Towards this ambition, the best practices offered in the NeOn methodology regarding future reuse, abstraction, and intelligibility, were particularly taken into account. This involved prioritising the literature review of existing semantic resources in the field, and either reusing, or aligning in some way, the developed model. This therefore led to a small number of alignments which were deemed as requirements for the ontology developed. Also, significant abstraction was stated as a requirement for the ontology, as this would allow its future alignment with upper ontologies or across to other domains with more ease. Intelligibility was also specified as a requirement, meaning that the ontology must make not only be logically consistent and valid, but somewhat intuitive for a trained person to understand. This soft requirement could be met by ensuring intuitive class hierarchies, avoiding very similar labelling of different entities, and excessive equivalence statements.

By achieving these goals, it was intended that the ontology would be more accessible, reusable, and modular, such that it could more easily contribute to the future development of a standardised semantic model for the domain.

#### 3.4.5.2 DEVELOPING A META-MODEL FROM REUSABLE ONTOLOGIES

Given a clear ontology scope, domain conceptualisation and an assortment of relevant knowledge modelling resources, these were then analysed for reuse potential and subsequently merged into a meta-model.  Of the broad list of resources identified in D2.1, those which were reused were the W3C semantic senor network (SSN) ontology and the socio-technical system (STS) ontology of van Dam [550]. The SSN ontology was adapted slightly in developing the meta-model to suit the WISDOM project, and the STS ontology directly reused, but only in part; again to suit the needs of the WISDOM project. This is subsequently discussed further. The guidance of the suggested upper merged ontology (SUMO) [551] was considered in merging these and developing the meta-model, so as to facilitate alignment with SUMO at a later stage.

#### 3.4.5.3 CANDIDATE ONTOLOGY CURATION

The WISDOM ontology's design extended the domain-independent meta-model to be domain specific, and was assisted by the automated web-based term-extraction process. The main stages in developing the draft ontology were to i) enumerate the domain's concepts, ii) build a class hierarchy from this which was aligned with the meta-model, then iii) define class relationships and iv) data properties before v) stating restrictions. The knowledge necessary for this task and the modelling decisions made was derived from the previous stage of scoping and knowledge acquisition and scenario identification, and was conducted alongside domain experts to further promote the ontology's accuracy and sufficiency. A full explanation of the nature of the various aspects of ontological modelling is beyond the scope of this thesis, but the main steps in are briefly summarised below.

##### 3.4.5.3.1 ENUMERATING CONCEPTS

The concept enumeration involved utilising the data objects produced previously in the task to list all of the types of object and any other key aspects of the domain's vocabulary. This involved a manual review of literature and domain expert communications as well as the IDEF0 models, use case models and scenarios; primarily for key nouns which represent classes of objects in the domain but also key adjectives and verbs which may represent properties and class relationships. The GIS data schema utilised by DCWW was also provided for analysis, which provided many key object types and properties. Finally, the automated web-based process offered a broad list of concepts used in the domain. This noun list was used to elaborate the concept enumeration and was used in the preliminary validation stage as an indication of ontology completeness.

##### 3.4.5.3.2 CLASS HIERARCHY DESIGN AND ALIGNMENT

From the list of relevant concepts, a list of domain specific classes was extracted, and this was iteratively developed into a hierarchy, aligned with the meta-model previously produced. The inherent modelling decisions were based on the knowledge acquisition conducted previously, common sense, and domain expert consultation. This produced a domain specific class hierarchy enriched with significant domain-independent abstraction to facilitate valuable inference and to better represent the underlying concepts in the domain, allowing its reuse or alignment with existing ontologies, upper ontologies and potential future applications. This class hierarchy represented a simple categorisation of types of concepts in the domain, including tangible and common sense knowledge such as ‘a waste water pipe is a type of pipe’ as well as more complex and abstract knowledge such as ‘a pipe is a type of physical arc and also a type of designed artefact’.

#### 3.4.5.3.3 OBJECT PROPERTY SLOTS

Object properties, which are relationships between instances of classes, were then modelled in agreement with the domain knowledge previously acquired. However, the object properties to be modelled were inherently considered during the class hierarchy design, as they are an integral part of the modelling decision process. This involved analysing the domain knowledge acquired and deducing statements such as ‘a sensor makes observations about a network entity’, then determining if the knowledge such a statement allows to be modelled is relevant and required within the requirement specifications given. Where the inferred statements were deemed necessary and valid, they were then formalised into the ontology. This resulted in many relationships between the social, sensing and physical systems in the domain, between the abstract and physical descriptions of the water network, between the designed and natural physical entities and between all entities and descriptive properties where they themselves were represented as concepts.

#### 3.4.5.3.4 DATA PROPERTY SLOTS

Relevant domain specific data properties were then added, to facilitate the ontology’s intended purpose of representing the current state of the water network. This implied that as well as static data such as pipe diameters and topologies, some dynamic data should be stored in the ontology, such as the latest reading from a sensor or the current flow rate in a pipe. This stage involved a thorough consideration of the software specification, so as not to duplicate the functionality offered by the event database whilst sufficiently capturing the data required about objects in the domain. The data property slots formalised are expected to be extended considerably as the WISDOM project matures and the data required by other software components evolves. Regarding the social modelling, much use was made of the previous work of capturing knowledge at the individual domestic consumer level through surveys to identify the key properties which describe residents from a water usage perspective.

### 3.4.5.3.5 RESTRICTIONS

Finally, restrictions (axioms) were modelled to supplement the potential inference and hence the new knowledge which may be inferred through the ontology. These included both ‘necessary’ and ‘necessary and sufficient’ restriction. Respectively these state that a class must adhere to a rule, and beyond that, if an individual adheres to the rule then it is a member of the class. This is useful in stating for example that a storage node must not change the water type between its input and output, and for stating that if an individual does change its water type from ‘raw’ to ‘potable’, it must be a treatment node, respectively. This rule development is related to the inference engine developed.

### 3.4.5.4 INSTANTIATING PILOT SITE KNOWLEDGE BASES

Pilot site knowledge bases were produced for each of the 3 Welsh sites and the Italian site. These have mainly been produced by reusing GIS data, sensor databases, EPANET [552] models, and social entity descriptions provided for the sites by industrial partners, as well as some manual data elicitation and input. The data reuse was accomplished in an automated manner using a Python application written as part of this task, for this specific purpose, following the manual federation of the data schema into the domain ontology as described previously. This data has been federated into RDF format and has been enriched through additional data gathered through domain knowledge and domain expert consultation. This utilised the RDFlib Python library [553] to store the RDF data in memory, and the Python CSV library to parse the input data, as well as manual pre-processing of the EPANET input file.

The knowledge bases were instantiated primarily by reusing GIS, sensor, and simulation data provided by industrial partners, through the development of a Python script which performed this conversion. This was then enriched manually with further sensor and social entity descriptions. The scripts extracted the relevant entities and properties from the GIS and EPANET files and sensor databases provided (following automated conversion to comma-separated value (CSV) format) and produced an RDF/XML serialization of the data aligned with the domain ontology. Specifically at the Welsh sites, as subtly different GIS schemas were provided for each pilot site, which needed to be converged to the same domain model, different scripts were made for each pilot site, although each followed a similar pattern. At the Italian site, an EPANET model was provided, the input file for which was split into relevant CSV files describing each type of entity, before utilizing the same Python approach as at the Welsh sites. The main functions of each part of these scripts are the same. This script is now briefly described, before the data enrichment through the process is described, and then each Abox is presented in turn. Each Abox file could be trivially merged with the Tbox if required, or kept separate until storage in the triple store.

The Python script to convert the GIS export files to RDF data contained 9 parts; each similar across the pilot site scripts, although sections 3-7 used different terminology and column placements, due to the different GIS exports. These 9 steps are now briefly summarized:

1. Import the RDFlib [553] and CSV libraries, to facilitate use of these file formats
2. Create an RDFlib graph object, and RDFlib namespace; the graph is used as a container for the triples and the Namespace variables shorten URIs
3. Use the Python CSV library to open the CSV file and create a reader object to parse the data into a list of lists
4. Iterate over each nested list, which each includes the data from a single row
5. Ignore the column header row, then print the new individual's name to the console
6. Create a new named individual using RDFlib, with a URI based on the base namespace, a pilot site and asset type string, and the entity's GUID from the GIS system
7. Create triples from all the relevant columns of the sheet to fully populate the description of the individual. The example only shows RDF:type and OWL:DatatypeProperty statements, but the actual code also automated object property statements.
8. State the size of the resultant Abox to the console.
9. Open an RDF file and write the data to it, using RDF/XML serialisation

By processing the GIS data, it was enriched alongside conversion into RDF format. In the GIS format the data is simply stored as a list of values and the meaning of the values is assumed by the software using the data. In the RDF/XML serialization the data is linked to properties which themselves have descriptions, and where appropriate the property itself is an object with its own set of object and data properties; therefore, the meaning is far more explicit and could be reused by other software with far less risk of misinterpretation.

The approach of extracting CSV files from an EPANET input file simplified the coding required, although an EPANET Python library could have been used to directly extract knowledge from the EPANET input file, which represents a possible avenue of future work.

Expressiveness and extensibility are significant benefits of the semantic modelling approach. For example, the ‘hasMaterial’ property is an object property which connects a pipe to a material, the material can then be described by properties such as surface roughness, for hydraulic modelling, or fracture toughness, for earthquake resilience simulation. These examples show respectively that the approach allows greater value to be derived from the initial data by formally describing it in a machine interpretable manner, and allows extensibility beyond its initial purpose with little effort. Further, semantic inference over the RDF form of the data allows greater value to be derived from the original data. For example the ‘goesToIpid’ is a datatype property which connects a pipe to an integer, but an SWRL is used to infer the knowledge that, given that the integer is the ID of another pipe, the latter is downstream of the former pipe. This is discussed more in the inference section.

#### 3.4.5.5 DEPLOYMENT AS A WEB SERVICE

The deployment of the ontology as a web service supports the benefits of a service-oriented architecture [15] and hence allows plug-and play capability with other software components of the WISDOM architecture. This software development was conducted in line with the software requirement specification produced for the ontology service. This was developed as a RESTful web service, written in Java, and based on the Apache Jena [554] suite of APIs for semantic web software development. The software was developed and tested on a local machine and has since been deployed on the secure cloud environment provided by ICL. This software is now at a mature stage where it is able to handle real-time SPARQL requests as well as custom functions for the most common foreseen uses of the ontology service, and meets the requirement specification. The ontology has been deployed through a web service to test its capability to handle real-time data, as discussed previously.

#### 3.4.5.6 ONTOLOGY VALIDATION

The ontology design process was followed by a collaborative and semi-automated preliminary validation, which was complemented by a secondary validation and revision iteration based on a meeting of industrial experts from relevant organisations. The initial validation consisted of using Protégé's [555] built in consistency checker, which determines if any of the explicit statements which have been made are contradictory. This was followed by competency question checking and iterative ontology revisions until the questions could be answered successfully, at this point the ontology could be assumed to sufficiently meet the stated objectives. However, it is critical to further test whether these stated objectives are an adequate representation of the intended objectives, and even further testing to determine whether these intended objectives are a valid representation of what the objectives should be. This testing has been conducted at a preliminary level through validation of the model by domain experts within the WISDOM project, in terms of: whether it meets their view of what the ontology should be, whether its statements are correct, and whether this is genuinely a valid representation of the wider domain. This was then further checked, still at a preliminary stage, through a semi-automated web-based term-extraction process, this involved referring to the work conducted at the requirement engineering stage, which is detailed later.

#### 3.4.5.7 DEVELOPMENT OF INFERENCE ENGINE

The inference engine used native OWL axioms as well as custom SWRL rules to infer domain specific and useful knowledge from that which is explicitly stated in the ontology, so as to produce a richer and more complete representation of each pilot site, at each time step of the ontology's instantiation, and to perform various checks which will raise alarm events if they fail. This followed a robust use-case based approach to elicit the requirements of the inference engine, and subsequently develop and test the inference in the context of real-world situations. These custom rules supplement the default inference capabilities available to all ontologies through reasoners such as the Jena native reasoner and the Pellet reasoner, with heuristic rules relevant to the management decision process. This work was tested in the Protégé software, and the capabilities and efficiency of the engine is reported.

#### 3.4.5.8 VALIDATION WITHIN WISDOM PLATFORM

Following the preliminary validation of the domain ontology, and the development of the other semantic components, these were tested together in the WISDOM platform as a web service, through its capability to sufficiently provide knowledge management services to the other platform components. This served to test and revise the previously produced requirement specifications based on the ontology service's contribution to the overall goals of the WISDOM project in supporting the other services. This formed an iterative process of semantic component revision and re-testing until the collective result was deemed satisfactory.

### 3.4.6 STAGE 3: GENERALISATION ACROSS DOMAINS

Based on the work conducted within the energy and water domains at the building and network scales, through the participatory studies, the hypothesis was finally explored in a more generic 'smart city' sense, to promote applicability to other domains and a wider range of usage patterns. This was intended to address the 4th research question, through a design science research approach. The ontologies from the 2nd stage of the investigation were unified through a smart city upper ontology, which builds on existing standards. A software framework was also developed to support the exploitation semantic models. This followed a design research approach, where design research in information systems aims to extend the state of the art in current system designs, whilst simultaneously producing knowledge which is useful to future practitioners, as discussed previously. This remit was well suited to answering the latter research questions, and especially question 4, where the proposed solution aimed to extend the state of the art, and learning from the development and testing of this next-generation software produced knowledge which would be useful to future practitioners. The stage was useful as it provided significantly more breadth and reusability to the proposed solution, and built on the learnings and artefacts produced through the 1st and 2nd stage.

The first aspect of this work was to scope the system to be developed and to produce a set of requirements to specify what was intended to be developed as a unification and extension of the previously conducted work. This involved reconsidering the work conducted in the 1st stage, especially the various scenarios developed, in light of the learning and outcomes of the 2nd stage. Specifically, the gap analysis conducted in stage 1 was revaluated, in terms of how much of the gap the work conducted had addressed. This also aimed to build on the learnings of the 2nd stage in terms of the limitations and reservations held about the systems engaged with and the outputs produced. After a scope of work for this stage had been produced in terms of its knowledge and software development, the main design and build process began.

Following the requirements engineering process, ontological modelling was conducted to facilitate the integration of data across domains. This was undertaken by developing an upper level smart city ontology, with a focus on utility networks, which domain ontologies could be aligned through. This built on relevant existing work in the domain as well as primarily aiming to integrate the ontologies developed previously, as well as emerging smart city standards. This began by reusing relevant ontological resources and extending these with concepts and relationships deemed necessary, based on the learning and experience gained previously. The previously curated ontologies were then aligned with the smart city ontology, to allow data to be contextualised across systems in a coherent manner. The ontology was then further extended to facilitate integration with a leading IoT data format, the Hypercat standard, based again on the learning and experience gained previously. This also aimed to harmonise the Hypercat standard with the ongoing W3C development of a ‘thing description model’. Following the curation of this ontological framework, it was tested through SPARQL query execution against the competency questions formalised.

A significant software development undertaking was then conducted to provide a coherent storage solution and suite of APIs for RDF, timeseries, BIM, CityGML, and Hypercat data. This work built separate servlets on an Apache Jetty server for each of the interfaces. SPARQL and Hypercat endpoints were built to interface directly with the triple store as a single point of truth. The triple store was also closely integrated with a KairosDB server to contextualise sensor data in a rich way. Integration with the BIM and CityGML servlets was loosely achieved, but was limited due to poor development and adoption of semantic web versions of the conceptual models underlying these standards. An API was developed for interfacing with BIM and CityGML models, based on open source Java libraries, but their data must be stored in their native formats rather than RDF. For the Hypercat servlet, no Java library was found, so one was developed to represent Hypercat objects internally.

Following the development of the software platform, it was tested against the requirements to evaluate its knowledge management capabilities, and its functionalities against the GAP analysis conducted. The system was then analysed, as well as its development process, which resulted in a number of lessons learnt and recommendations for future practitioners. This knowledge, alongside the resulting artefacts, were the primary outputs of this stage's design research based process.

## 3.5 RESEARCH EVALUATION ISSUES

### 3.5.1 THEORY OF RESEARCH QUALITY EVALUATION

As a multi-stage research design, each stage utilised a separate validation approach, nested within an overall strategy to promote the validity of the investigation and accuracy of the conclusions drawn. The approach aimed to ensure that the conclusions drawn were well reasoned and that the processes undertaken were appropriate, well considered, and accomplished with integrity. Further, bias was mitigated by seeking expert review of the work at various stages, including peer review prior to several publications. As the majority of the research was qualitative in nature, the conclusions drawn inherently incorporated subjectivity in the interpretation of the evidence, and the participatory nature of the 2$^{nd}$ stage also led to researcher bias potentially affecting the data collection and analysis process. However, this is a common feature of all participatory research, action research, and design research, which are all major methodological approaches in information system research, and so doesn't necessarily invalidate the research outcomes, provided that these factors are well-considered in the design of the research and analysis. Given this, criticisms (and the mitigation of these) around the research quality should be framed with an emphasis on assuring *trustworthiness* rather than *validation* in the true sense of the latter word [536], [556], as this applies a positivist perspective, which is broadly considered unsuitable for research outside of natural sciences. This trustworthiness has been said to consist of credibility, transferability, dependability (repeatability), and confirmability (neutrality) [556], with methods of promoting these aspects summarised in Table 10. However, it is important to recognise that this school of thought is arguably drawn from the period before ‘modern’ information system research and computer science research, and so is primarily intended for studying social systems, rather than computer systems and human-machine interactions, so the modern context must be considered in utilising traditional qualitative evaluation techniques. For example, [536] states in a more modern work that action research as a whole requires its own set of quality criteria, distinct from more traditional social science.

**Table 10: Means of promoting qualitative research quality [556]**

| Trustworthiness aspect | Methods for promoting aspect |
|---|---|
| Credibility | Prolonged engagement, persistent observation, triangulation, peer debriefing, negative case analysis, referential adequacy, member-checking |
| Transferability | Thick description |
| Dependability | Inquiry audit |
| Confirmability | Confirmability audit, audit trail, triangulation, reflexivity |

Qualitative and quantitative research need significantly different validation approaches. The majority of the research design presented utilised a qualitative approach. However, some quantitative research was also required to fully explore the research questions, such as the performance speeds of the developed artefacts compared to current alternatives. Therefore, the validation of each stage needed to account for this slightly mixed methods approach.

### 3.5.2 APPROACH TO RESEARCH QUALITY ASSURANCE

In the first stage, the theoretical study was used primarily as background research and to scope the remaining project, and so carried less onerous validity challenges. However it was important to ensure that the primary outcomes of this stage were suitable for use by the remainder of the project, and could serve as the primary evidence base for answering the first research question. This involved the review of the impact scenarios by domain experts, as well as the peer review of aspects of the literature review. Also, the use of a comprehensive literature review promoted the validity of the remaining work, by well grounding it in the existing solution space. Two main relevant types of literature review are defined: narrative reviews and systematic reviews, although a meta-synthesis of literature would be somewhat relevant [557]. The primary difference between a narrative review and a systematic review, is the clearly defined methodology for conducting the review, in terms of defining a research question, choosing literature and analysing the literature to answer the question. In this manner the theoretical study aimed to promote a rigorous and unbiased approach, where the criteria for the aforementioned aspects were presented in section 3.4.3. Following from the initial literature review, alongside the participatory research projects, impact scenario development envisioned a ‘possible case’ of how emerging technologies could assist in overcoming challenges reported by people, organisations, and nations. This then formed the possible future point of a GAP analysis, which compared the current state observed in industry and literature against the ideal envisioned future, and considered a possible pathway towards this by overcoming specific limitations of the current state.

In the second stage; the participatory action research stage, validation was in part achieved by the participation alongside experts, as well as the rigorous iterative process adopted. Building on the seminal work of Lincoln and Guba [556], Williams [558] outlines the aspects of qualitative rigour pertinent in information systems research, highlighting the necessity to consider rigour against the initial research goals. She also identifies aspects which contribute to quality action research: clear descriptions of the processes, detailed documentation of each stage, use of complementary interpretation techniques such as triangulation, and statements of the research philosophy, objectives, and *a priori* knowledge. The latter of these has been provided in this chapter, and the former are provided within the relevant later chapters. Champion and Stowell [559] concur with this sentiment that validating action research is not a clearly defined issue, but state that it requires a thorough consideration of the context, and a rigorous and documented approach, directly towards solving the research questions. They propose that the traits required are: documentation of participant selection, engagement mechanisms, authorities over actions and processes, relationships, and learning outcomes. Again, these are documented within this chapter and across subsequent chapters. Herr and Anderson [536] offer a comprehensive and modern description of best practices in preparing an action research dissertation, including a thorough consideration of positionality, and also of the quality criteria for action research, the latter of which is summarised in Table 11, and broadly guided the research design.

**Table 11: Quality criteria promoting each action research goal [536]**

| Action Research Goal | Quality/Validity Criteria |
|---|---|
| 1) Novel knowledge contribution | Dialogic and process validity |
| 2) Action-oriented outcomes | Outcome validity |
| 3) Researcher and participant learnings | Catalytic validity |
| 4) Local result relevance | Democratic validity |
| 5) Sound and appropriate methodology | Process validity |

The quality assessment of action research is most relevant in answering research question 3, and to a lesser extent 2 and 4, as question 3 aimed to understand the value of SWoT systems, which is inherently an investigation of a socio-technical system. Specifically, this aimed to determine if there was value to persons and organisations involved in developing software, or using this software to make decisions within operational smart city systems, or whether there was value in overcoming challenges within the wider STEEP contexts of cities. Questions 2 and 4 then emphasised more the problem-solving and action-oriented goals of action research, in determining how the state of the art of system and component designs should be extended to overcome the observed technological gap(s) and providing knowledge to future practitioners from this. By following the guidance of these sources, the research design aimed to embody best practice in eliciting high quality, valid, learnings from the participatory action research work conducted.

Finally, in the third stage, a design research approach was conducted to unify and extend the outcomes of the 2$^{nd}$ stage, which included significant aspects of design science. This involved designing a generic framework, which was tested and refined through experimental methods to evaluate various functional arrangements of components, towards overcoming the wider gap identified in smart city ICT systems. Whilst design science inherently focuses a significant amount of effort on the creation of an artefact, the research outcome of design science is rather the knowledge *about* the artefact(s) [560]. Design science has also been defined as a science which constructs and outputs novel meta-artefacts from a pragmatism-oriented epistemological perspective [541], and is increasingly relevant in information system research [542], [546]. It has been stated that 4 aspects are required to validate the outcomes of a design research process: artefact success, generalisation, novelty, and explanation capability [560]. The first of these, artefact success, is said to relate to the efficacy and efficiency of the artefact for its intended purpose, which echoes the 'validation square' of Seepersad et al. [561], which emphasises only these two aspects. The artefacts' successes were determined based on their specific design requirements, specified for each in subsequent chapters. The generalisation capability of the knowledge were promoted through stage 3, which tested this in new domains and use cases, and was further considered through consultation with experts and the literature review. The novelty of the research is derived from the thorough literature review and consultation with experts. The explanation capability was derived from the researcher learning which occurred throughout the process, the triangulation approach to testing, and the deep experience gained through extensive engagement with the participatory projects.

The validation also considered the suitability of technology choices for each functional component, and compared usage patterns of semantics in supporting the ICT traversal of the reference model. These tests also served to compare the framework and the use of semantics with perceived traditional approaches such as relational databases, and suggested how they could be used to complement each other. Whilst the 3$^{rd}$ stage followed a 'pure' design science research approach, there were significant elements of design science within the participatory projects of the 2$^{nd}$ stage, and so these considerations are relevant also to the evidence used from those towards answering the research questions.

From a design science research perspective related to the artefacts produced in both the 2$^{nd}$ and 3$^{rd}$ stages, the qualitative aspects of the software developed (such as its functional capabilities) were validated as an extension of the state of the art by analysing them thoroughly against examples from the literature, and expert consultation. The quantitative aspects of the software developed were validated through direct comparison with alternative approaches, or against benchmarks, or against evidence from the literature. Finally, the ontologies themselves were validated as accurate and sufficient domain representations by domain experts themselves, a thorough process of automated ontology checking, and competency question checking through SPARQL queries. Further details and evidence of these processes are presented in the subsequent chapters.

## 3.6 ETHICAL ISSUES

Within a research project, ethical considerations must be put in place to morally ensure that professional, legal and social obligations are adhered to [545]. Computer science research can engage with sensitive data and system functions with direct and indirect impacts on personal and organisational interests and safety, and can dramatically affect humanity's relationship with the world and each other [562]. Ethical decisions in ICT are primarily influenced by 3 aspects: i) the researcher's moral code, ii) informal ethical code in the researcher's environment, and iii) exposure to a formal code of ethics [563]. In order to ensure such obligations were adhered to, the researcher followed the formal ethical guidelines of Cardiff University [564], within the BRE Institute of Sustainable Engineering research group, a respected and professional organisation. The work did not utilise personal data, didn't conduct personal interventions, and didn't directly control target systems in vivo, which significantly reduced the burden of ethical consideration. The main aspects requiring ethical professionalism were the security and privacy of sensitive organisational data, suggesting actions to decision makers, and dissemination of learnings from participatory action research projects, as well as considering the nature of the work in the broader context of computer science and the current socio-politico-economic landscape.

## 3.7 SUMMARY

This chapter has provided a description of the methodology and research design, as well as the philosophical groundings of the approach adopted. The chapter began by stating that the research has adopted the paradigm of pragmatism, and explained the reasoning behind this decision. The high-level research approach was then described as a combination of participatory action research and design research, in a multi-stage methodology which utilised some quantitative methods in a primarily qualitative mixed-methods approach. This research design was then described and justified in detail as a sequence of 3 stages in line with the research questions posed. This 3 stage approach was chosen to provide a systematic means to gather a broad evidence base, which iteratively targeted a more specific in-depth consideration of the hypothesis.

The first stage was described as a theoretical study to robustly situate the work within the existing solution space, and to provide a well-reasoned direction for the subsequent stages. The main processes in this stage were a literature review, and a scoping task. This resulted in a clear statement of the gaps in the literature, what would constitute a significant step towards filling these, and a set of impact scenarios as to the potential pathways for the considered technologies to assist in smart cities.

The second stage was then a participatory research process, which shifted from participation to direct intervention over the course of engaging with 6 research projects. This provided a systematic means to build on the theoretical study in real world systems. Initially prioritising the observation of existing practices and expert involvement helped with the mitigation of researcher bias and supported a pathway towards genuine insight. By firstly observing the role of ontologies at a higher level, this guided the later project engagements, where the lessons learnt were built on in directly developing and testing ontology-driven IoT systems.

The third stage then unified the distinct threads of research from the second stage into an overarching architecture for application layer interoperability in IoT systems. This primarily involved two aspects; the unification of the separate knowledge modelling work conducted for each project, and the unification of the software development work conducted for each project. The former of these was achieved through ontological alignment and abstraction through a higher level ontology. The latter of these involved leveraging the lessons learned across the projects to produce a generic platform which is sufficiently flexible to meet a wide range of smart city use cases.

The proposed research design was developed in alignment with the previously stated hypothesis and research questions; the mapping between these is as follows:

- The 1st research question, regarding theoretical underpinnings, including challenges, scenarios and requirements, was answered primarily by the 1$^{st}$ stage, with supporting evidence from the participatory research projects.
- The 2$^{nd}$ research question, regarding how semantic web technologies can integrate IoT and AI to meet these requirements, was primarily answered across the 2$^{nd}$ stage of the research
- The 3$^{rd}$ research question, regarding the value of adopting such an approach, was answered again through the 2$^{nd}$ stage of the research, with supporting evidence from the 3$^{rd}$ stage.

- The 4th research question, regarding extensibility and generalisation capability of the artefacts and knowledge, was primarily answered through the 3rd stage, with supporting evidence from the 2nd stage.

By following the investigatory process described in this chapter, several iterative learning cycles were conducted, which each produced a number of software and intellectual outputs. The following chapter describes the details of the systems developed and the outputs of their development and testing processes, after first presenting the theoretical framework and high-level scenarios developed.

# 4 OUTPUTS AND RESULTS

This chapter presents the results of the three stages of experimental methods described in the methodology chapter, starting with the high level scoping and requirements analysis conducted through scenario development. This is followed by the results of the iterative action research cycles conducted through the second stage of the methodology within case studies in energy and water systems. Finally, the results of the third stage are presented, where a generic smart city framework was developed to unify the previous avenues of research and software development.

## 4.1 THEORETICAL STUDY AND SCOPING

This section presents the outcomes of the first stage and early aspects of the participatory research of the methodology proposed, where the main outcomes are impact scenario identification, use cases for semantic technologies, and aspects of requirements engineering for the overarching project. This builds on the research gap identified in the literature review towards an understanding of the potential and requirements of semantic technologies, in line with the first research question. This aims to act as a common framework and reference point for the action research iterations undertaken through the second stage, which the third stage then uses to unify these separate threads, and to promote extensibility to broader smart domains. The first subsection provides an overview and conceptual framework of the impact scenarios envisaged in smart city systems of systems. The second and third subsections then present in further detail the use cases and scenario descriptions for the energy and water domains considered in depth in the second stage, at the intra-domain level of complexity. The final subsection then proposes use cases which extend the role of semantic technologies more to the inter-domain level of complexity, where greater heterogeneity is experienced, to guide the third stage of the overall investigation. It should be noted that the literature review itself formed part of the 1st stage of the investigation.

### 4.1.1 SYSTEM OF SYSTEMS CONCEPTUAL FRAMEWORK

This section outlines a high-level conceptualisation of impact scenarios within and across smart domains, following a system of systems approach to smart cities.

Specifically, instances where the integration of IoT and machine intelligence may deliver value are proposed, and then briefly expanded upon. Three levels of detail were assumed in proposing impact scenarios: i) within an organisation, ii) across organisations within a domain, and iii) across organisations across domains. These distinctions were chosen as they represent levels of semantic heterogeneity and socio-political barriers for data exchange. The levels are not intended to be definitive or prescriptive, but were observed through the investigation and serve the purpose well. Figure 23 illustrates this concept by showing the hierarchical socio-technical context of water systems. In the current landscape, each node in Figure 23 is mostly isolated from the others from an automated technological perspective. For example, a city could benefit significantly if the energy and water domains were optimised alongside each other so as to mitigate pain points in both systems.

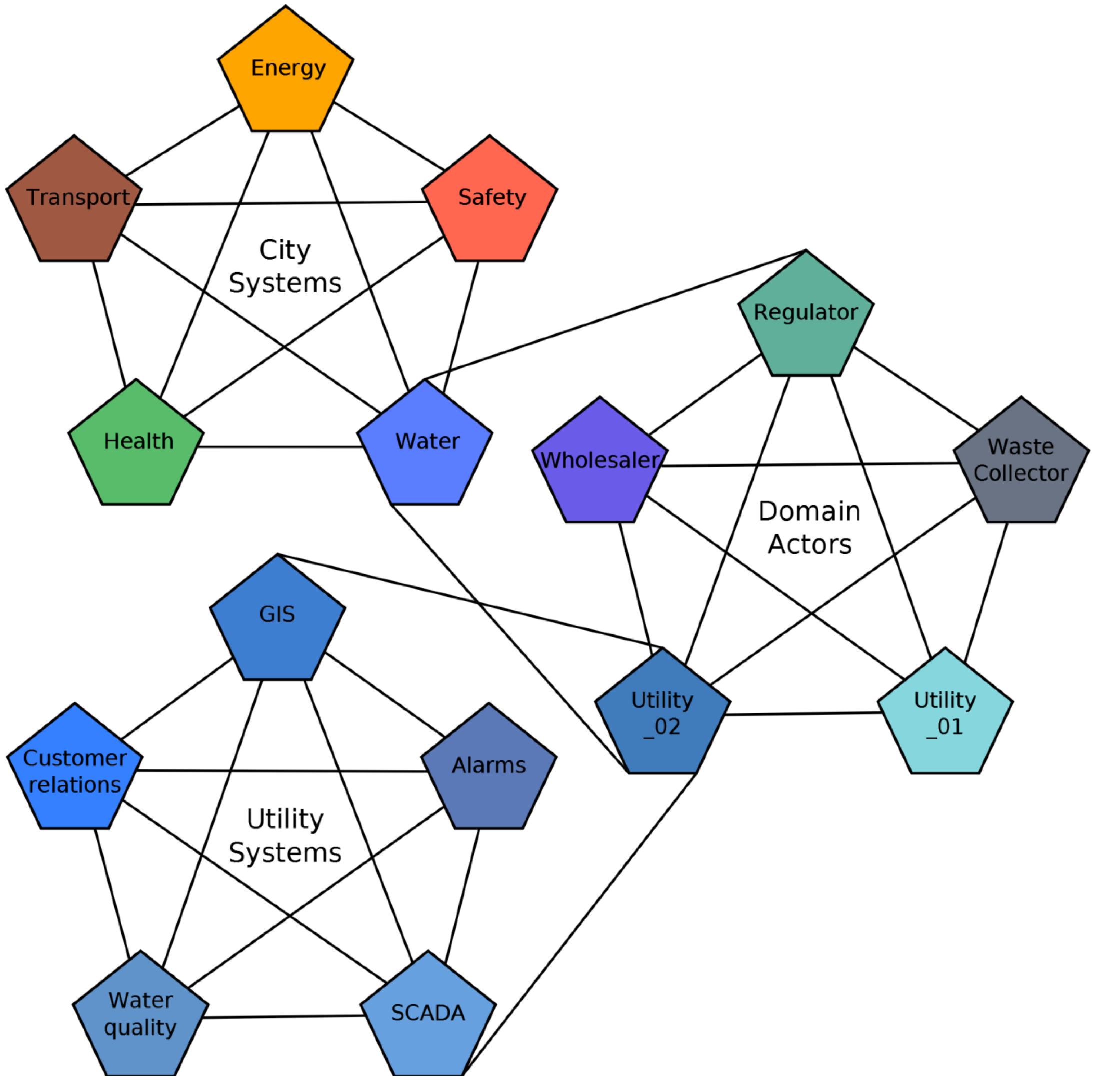


**Figure 24: Illustration of system of systems nature of smart city domains as three levels of detail (not exhaustive)**

At the lowest level of detail, across a business's systems, less legal and socio-political barriers exist in relation to data sharing, broadly speaking. These systems are typically bounded by the vendors which developed the component, the business area which the system provides value to, or the group of staff which use the software. Whilst significant heterogeneity exists (such as between OT and IT systems), the terminology used and data semantics are likely to be closer together than at higher levels. Scenarios which integrate automated consideration of the decision space across these systems aim to aid the owning organisation, which are typically aligned with higher level goals through regulation and sometimes through market forces.

At the middle level of detail, there are clear boundaries between social entities where system integration does not occur, although this is often not due to technological issues, such as in sharing data between one utility's GIS system with another utility's GIS system. There are often deeply embedded semantic differences between organisations which consider the same system from different viewpoints, partially due to the lack of standardisation in this area, and partially due to the different priorities of each organisation in utilising the system. Scenarios at this level must be agreeable to the stakeholders involved, and must bridge the syntactic, semantic, security, and trust aspects of interoperability between their respective systems.

At the broadest level of detail, domains can be generally conceptualised, but the boundaries between these in terms of their actors are less well-defined than between the systems which belong to each stakeholder, as higher level actors may have roles across these higher level domains. Broadly speaking, these domains are bounded by the service or value they provide to society, such as energy, water, or healthcare. Traditionally, these would represent separate value chains from production of raw materials to delivery of product or service to customer. However, integrating aspects of operational decisions across these domains may provide significant value in terms of economic, societal, or environmental performance indicators.

It is proposed that value could be derived by many decision makers, both in normal operating conditions and in mitigating the impact of a disruptive event, by developing software which can integrate data and services across the silos such as those identified, both at design time and automatically at runtime. A difference is noted in developing information systems depending on whether the heterogeneity of the data is known at design time or not, especially when advanced applications such as machine intelligence must understand the context of the data.

The energy domain represents a significant opportunity to test the hypothesis in a system which strongly impacts in social, economic, and environmental terms. Following on from this, the water domain represents a natural progression to evaluate the hypothesis across domains, as the potential of the water-energy nexus has been recognised and both are utility-centric sectors. Therefore the following sections explore scenarios within these domains, and finally across them.

### 4.1.2 THE BUSINESS INTELLIGENCE IOT STACK

A theoretical consideration was present in the 1st stage of the investigation, through the literature review, as well as in the 2nd stage through engagement with a broad range of experts. From a conceptual perspective, this produced an understanding of the value proposition of semantic technologies in IoT systems, which is to address the disconnect between the emerging IoT technological stack and the business intelligence pyramid, as illustrated in Figure 24.

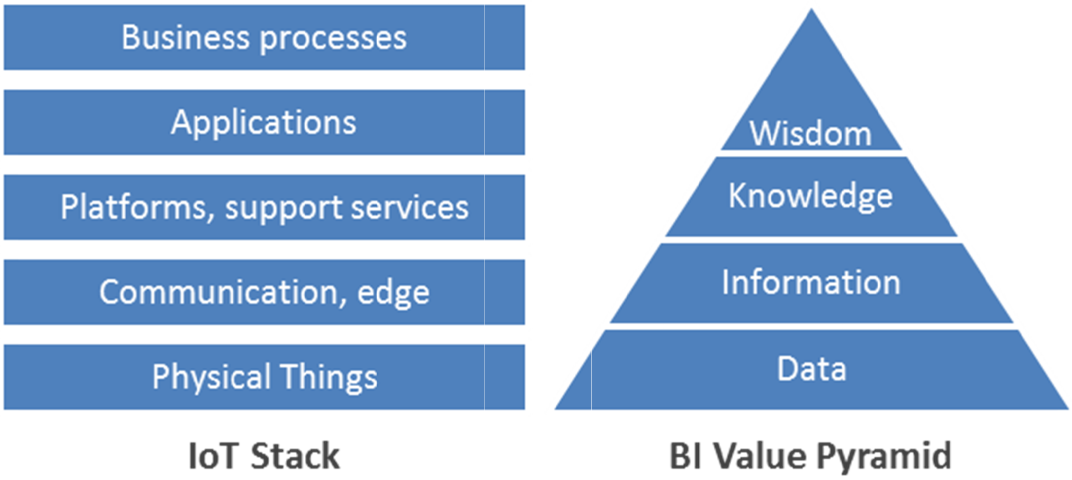


**Figure 25: Contrasting perspectives of IoT and BI**

Specifically, whilst the traversal of the rungs of the BI value pyramid is well understood in traditional, siloed, enterprise systems, it is not yet well understood how to translate this to disruptive IoT technologies. However, through the work conducted, it became apparent that semantic technologies can play a key role in this process, which resulted in the conceptual state model illustrated in Figure 25 and introduced below.

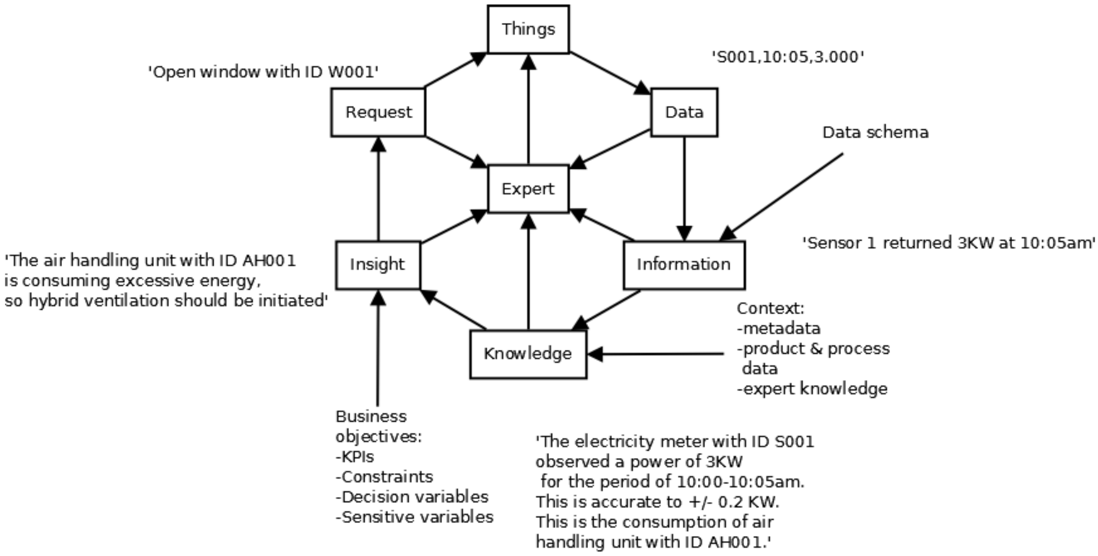


**Figure 26: Reference model for leveraging BI theory in IoT systems through semantic technologies**

Figure 25 shows a cycle around the perimeter which mirrors concepts in BI theory, showing the evolution of the value of knowledge through ICT. The model also includes a *'Request'* state, whereby the optimal actuation is eventually determined by the system after considering the entirety of the data's context. An expert person is placed in the middle of the cycle, and arrows indicate that the human can short-circuit the machine process at any time. This allows the model to be generic, and applied in different context, by adapting the point at which the expert typically resumes control. The arrows outside the cycle show the required semantic context which must be added to the data for it to progress to the next stage, and the comments around the diagram are example instances of each state.

The model was not used prescriptively, but was conceptualised iteratively throughout the investigation, and was used as a reference point for the role of semantics in the target systems engaged with. It is critical to note that at each stage in the cycle, multiple streams are likely to converge to produce the next stage. For example, several pieces of data are needed to create useful information, and several pieces of information are needed to create useful knowledge. As well as providing business and systemic context, this interoperability is critical to the role of semantic technologies.

### 4.1.3 ENERGY DOMAIN SEMANTIC TECHNOLOGY USE CASES

Following pilot site and business process analysis, this section proposes impact scenarios which test the role of semantic technologies in delivering value by integrating IoT and AI, to guide the work conducted in the second stage of the overall investigation. From the most granular level of detail to the most complex, the proposed scenarios are: i) building optimisation for energy minimisation, ii) smart home prosumer optimisation, iii) aggregation of domestic flexibility for grid stability and prosumer ROI, and iv) optimisation of a polygeneration microgrid. These scenarios are described in terms of their value propositions, impact pathways and pertinent data in Table 12 - Table 15.

**Table 12: Impact scenario description for E01: building energy optimisation across systems**

| ID#: E01 | Building Energy Optimisation Across Systems |
|---|---|
| Value Proposition | Reduce energy consumption |
| Description | This scenario aims to provide a more holistic solution to the pilot building's energy management by integrating temperature and occupancy data with simulation and optimisation to control the window actuation, air handling units, shading system, temperature setpoint for radiators, and lighting. |
| Independent variables | Lighting state (on/off)<br>Heating temperature setpoints (18°C-26°C)<br>Atrium roof window opening temperature set-point (16°C-24°C)<br>Shading state of each shade (on/off) |
| Input data | From facility management company:<br>Occupancy<br>Wind speed<br>Rain intensity<br>Current window state (open or closed)<br>PMV (comfort predictor from simulation) (-0.5 to +0.5)<br>Room temperatures<br>Solar radiation on external surfaces (E/S/W) |
| Pilot site | Residential care home 'FORUM building' in Eindhoven, The Netherlands |

**Table 13: Impact scenario description for E02: domestic prosumer optimisation**

| ID: E02 | Domestic Prosumer Optimisation |
|---|---|
| Value Proposition | Improve ROI for DG, storage, and load flexibility, and improve reliance on renewables |
| Description | Within individual households with various combinations of micro-renewables, energy storage, and flexible loads, the integration of weather and demand predictions, system descriptions, and market descriptions (such as variable tariffs), and artificial intelligence technologies, can optimise the operation of the smart home to the preferences of the home owner. This will primarily involve maximising the reliance on micro-generation, reducing the amount of peak-tariff energy consumed from the grid, and respecting the owner's desired flexibility of loads such as through washing machine cycle finish deadline. |
| Independent variables | Flexible load schedules (deferment, curtailment)<br>Battery charge/discharge state<br>Buying/selling from/to grid |
| Input data | From the home owner:<br>Battery level<br>System descriptions<br>Flexibility preferences (deadlines, min temperatures, min luminance)<br>Smart metering data<br>Demand predictions<br>From external sources:<br>Weather data and predictions<br>From grid operator/utility:<br>Market descriptions (tariffs for buying/selling from/to grid) |
| Pilot site | |

**Table 14: Impact scenario description for E03: prosumer aggregation**

| ID: E03 | Prosumer Aggregation |
|---|---|
| Value Proposition | Improve average demand profile of area to reduce peak and total load on grid, improve stability and hence confidence in demand profile, improve ROI for prosumers, improve reliance on renewables, and reduced transmission distance. |
| Description | By encapsulating the 'domestic prosumer optimisation' intelligence within a software agent and introducing many of these to aggregation |

| | |
|---|---|
| | agents and a grid agent, which varying the energy tariff dynamically and offer a market for domestic flexibility, the overall performance of the district can be improved. |
| Independent variables | To domestic agent:<br>Flexible load schedules (deferment, curtailment)<br>Battery charge/discharge state<br>Buying/selling from/to grid<br>To aggregator agent:<br>Flexibility tariff<br>Homes within portfolio<br>To grid agent:<br>Flexibility tariff<br>Energy tariff |
| Input data | As with 'domestic prosumer optimisation', plus:<br>Congestion points in network<br>Grid stress/capacity |
| Pilot site | |

**Table 15: Impact scenario description for E04: polygeneration microgrid optimisation**

| ID: E03 | Polygeneration Microgrid Optimisation |
|---|---|
| Value Proposition | Improved ROI for microgrid operator, improved reliance on renewables, improved service delivery for facility managers, improved demand profile for grid operator, more confidence in grid demand profile, reduced transmission distance. |
| Description | Demand profiles will be predicted for the microgrid's 5 public buildings through data mining or simulation, which will then be integrated with dynamic data and system descriptions. This will allow machine intelligence to optimise the control of the microgrid's energy hub which produces power and heat for the district. Specifically it will allow the optimal balance between gas boiler, gas CHP, biomass boiler, heat storage, and grid buying/selling to be pursued so as to minimise costs whilst also reducing the environmental impact of the energy production. |
| Independent variables | Generation mix setpoint schedule (plant start up/shutdown times and output rates)<br>Balance between heat and power production from CHP<br>Amount of energy bought from/sold to grid<br>District heating network supply temperature and flow rate<br>Energy storage charging/discharging rate |

| | |
|---|---|
| Input data | From the energy hub operator:<br>DHN and microgrid descriptions (topologies, materials, efficiencies if known)<br>Real-time metering data<br>Energy hub description (plant efficiencies, start up times etc.)<br>From the grid operator:<br>Variable tariff details<br>From the building management companies:<br>Metering data (heat and power)<br>Building descriptions (CAD/BIM as available)<br>From external source:<br>Weather data and predictions |
| Pilot site | Ebbw Vale |

#### 4.1.4 WATER DOMAIN SEMANTIC TECHNOLOGY USE CASES

As with the energy domain, and following the lessons learned therein, a number of impact scenarios were devised where semantic technologies may unlock value in the water domain by integrating IoT and AI in innovative ways. These are presented in Table 16 - Table 18, again in terms of their value proposition, impact pathway, and pertinent data.

**Table 16: Impact scenario description for W01: water utility integrated network monitoring**

| ID: W01 | Water Utility Integrated Network Monitoring |
|---|---|
| Value Proposition | Improved business processes from integration of data sources in a manner which supports insight extraction. |
| Description | Integrating existing GIS and SCADA systems will enrich of existing decision support systems with data regarding compliance monitoring, network malfunctions and water quality indicators etc. A better understanding of the current state of the water network will allow the observation of spatio-temporal trends, a more rapid response time and an improvement of business processes. |
| Independent variables | N/A |
| Input data | From water utility:<br>Water flow rates<br>Water pressures<br>Valve states<br>Pump states<br>Sensor health states |

| | |
|---|---|
| | pH<br>Stage height<br>Temperature<br>Chlorine residual<br>Rainfall<br>Salinity<br>CSO spill (binary)<br>Water density<br>Water turbidity<br>Asset locations<br>Pipe locations<br>Asset & pipe descriptions (material, type, size etc.) |
| Pilot site | Tywyn & Aberdovey, Cardiff, Gowerton, La Spezia |

**Table 17: Impact scenario description for W02: demand optimised management**

| ID: W02 | Demand-Optimised Management |
|---|---|
| Value Proposition | Matching the availability of water to the demand for water should reduce energy consumption, non-revenue water (leakage and evaporation), and maintenance costs, and reduce the number of alarms by reducing the strain on the network. |
| Description | The integration of predictive models, optimisation algorithms and decision support tools will allow the suggestion of set point schedules and resource management schemes for water network operational assets. Specifically, simulation and optimisation models will be integrated with weather data and predictions to suggest options to functional managers and information regarding the implications of control strategies. A multi-objective optimisation will reduce the amount of pumping required and the peak pressures, hence reducing leakage and energy cost. Reducing the time water resides in reservoirs will also reduce evaporation through more of a just-in-time approach. |
| Independent variables | Reservoir set points and schedules<br>Pump set points and schedules<br>Pressure reducing valve actuation<br>Control valve actuation |
| Input data | As with 'water utility integrated monitoring', plus:<br>From external source: weather data and predictions<br>From consumers: smart metering data<br>From Environment Agency: river levels and rainfall gauges<br>From water utility: Sensor descriptions |

| | |
|---|---|
| Pilot site | Tywyn & Aberdovey |

**Table 18: Impact scenario description for W03: predictive fault impact support**

| ID: W03 | Predictive Fault Impact Support |
|---|---|
| Value Proposition | Improved regulatory compliance from rapid fault detection and correction, improved customer relations from pre-emptive communication and quicker fault correction. |
| Description | By integrating the static and dynamic data with a semantic reasoning engine, it is possible to produce more insight about a problem in the water network which may have caused an alarm, and possible solutions. By integrating this with GIS and customer data it is possible to infer which customers may be affected, to gather more data about the issue and to pre-emptively inform them of the problem. |
| Independent variables | N/A |
| Input data | As with 'water utility integrated monitoring', plus:<br>From customer database: Customer names, contact details, and property locations<br>From maintenance personnel database: Staff locations, capabilities, contact details<br>From water utility: alarm data |
| Pilot site | Tywyn & Aberdovey, Cardiff, Gowerton |

### 4.1.5 ENERGY-WATER NEXUS USE CASES

Following the delivery of successful outputs in the energy and water domains, a unifying platform was made which aimed to achieve similar impact pathways by integrating the energy and water domains at various scales. This was based on several conceived scenarios where integrating IoT and AI across these systems may bring value, which are presented in Table 19 -Table 21, again in terms of their value proposition, impact pathways, and pertinent data. These scenarios aimed to test the hypothesis at the higher level of complexity and heterogeneity, across domains, from which extrapolations could be considered into the value of the approach in other smart domains.

**Table 19: Impact scenario for EW01: water utility load-shifting**

| ID: EW01 | Water Utility Load-Shifting |
|---|---|
| Value Proposition | Reduced energy cost for water utility, improved demand profile for energy utility, and improved confidence in demand profile. |
| Description | As water utilities are significant energy consumers, shifting portions of their loads to off-peak times would benefit grid operators significantly. By integrating the 'demand-optimised management' scenario as a software agent with a grid agent which proposes variable pricing based on predicted load schedules and grid capacity, the water-energy nexus may be optimised for joint benefit. |
| Independent variables | To water network agent:<br>Load shifting (pump start/stop times)<br>To grid agents:<br>Variable tariff |
| Input data | As with 'demand optimised management' scenario, plus:<br>From water utility:<br>Energy consumption data and predictions<br>From grid operator:<br>Predicted load schedules<br>Grid capacity |
| Pilot site | None |

**Table 20: Impact scenario description for EW02: water utility integration with distributed generation**

| ID: EW02 | Water Utility Integration with Distributed Generation |
|---|---|
| Value Proposition | Reduced energy cost for water utility, improved reliance on renewables, reduced transmission distance, improved demand profile for grid operator, improved ROI for local energy producers, reduced energy bill for local consumers. |
| Description | As an extension of the prosumer aggregation scenario, following investment by either a water utility or local prosumers in renewables or energy storage, it could be conceived that the utility's asset may interact with a low-voltage grid. This would improve the ROI for the utility investing in renewables (such as for a pumping station) as it |

| | |
|---|---|
| | would be able to sell surplus energy to local prosumers at off-peak times, and purchase their excess energy at a cost lower than from the grid. By protecting each actor's interests through multi-agent based control, a situation may emerge which is beneficial to all parties compared to not integrating the utility's renewables in the local grid. |
| Independent variables | As with 'prosumer aggregation scenario', plus:<br>To water asset agent:<br>Load shifting (pump start/stop times)<br>Energy exchange with grid<br>Energy exchange with local aggregators<br>Energy storage charge/discharge rate |
| Input data | As with 'prosumer aggregation scenario', plus:<br>From water utility:<br>Any flexibility in demand profile<br>Device and system descriptions at asset<br>Demand predictions and smart metering data at asset |
| Pilot site | None |

**Table 21: Impact scenario description for EW03: integrating smart home energy and water demand-side management**

| ID: EW03 | Integrating Smart Home Energy and Water Demand-Side Management |
|---|---|
| Value Proposition | Reduced water and energy bills for consumer, improved demand profiles for water and energy utilities, improved convenience for smart home owner. |
| Description | Given the likely future situation of both energy and water dynamic pricing, it would be beneficial to home owners for load scheduling to optimise across both of these systems to minimise their costs, and their environmental footprint if desired. This will require the integration of semantics across these systems at the domestic level and coordination of devices which impact both variable tariffs at the same time, such as washing machines. By extending the 'domestic prosumer optimisation' scenario to include water tariffs too, this added complexity may be accounted for. This could be further extended to include the 'demand optimised management' and 'prosumer aggregation' scenarios for added benefits to the water and |

| | |
|---|---|
| | energy utilities through adaptive pricing and ultimately optimise synergistically across both systems. |
| Independent variables | As with 'domestic prosumer optimisation' scenario, plus:<br>Water demand flexibility (deferment/curtailment) including 'hard' scheduling (e.g. washing machine) and 'soft' scheduling (e.g. bath time & duration) |
| Input data | As with 'domestic prosumer optimisation' scenario, plus:<br>From the home owner:<br>Water consumption details of devices & functions<br>System descriptions<br>Consumer objectives (i.e. purely financial or some bias towards environment)<br>Water meter data<br>Demand predictions |
| Pilot site | None |

## 4.2 ENERGY SECTOR INVESTIGATION

This section presents the results of the participatory action research undertaken in the energy domain through 2 projects, firstly at the building level, and secondly beyond the building level to the local grid level. Each project is introduced, as well as the role of the project within the present investigation, before presenting the work conducted and artefacts developed within the project, and finally project results. Note that the contribution of the investigator to each project is detailed in the methodology section.

### 4.2.1 SMART BUILDINGS

#### 4.2.1.1 OVERVIEW & PROJECT DESCRIPTION

The first iteration of action research was conducted in the energy domain at the building level. This involved engaging in and contributing to an ICT solution which achieved semantically-enabled advanced application decision support. The work was conducted in the context of the EC FP7 project entitled ‘Knowledge-based, holistic, energy management of public buildings’ (KnoHolEM).

This project involved the installation of a relatively small number of wireless sensors into pilot buildings, to detect properties such as temperature, air speed, occupancy, and humidity. The data from these was then communicated to a cloud-based retrofit building energy management system developed within the project. The solution was primarily tested within a mixed mode residential care home in the Netherlands. The primary goal of the Project was to produce a BEMS for retrofit into public buildings with minimal investment, to exploit the enhanced sensing infrastructure and any existing BEMS, augmented with analytics and visualization components through a semantic web approach.

This involved a semantic knowledge base, which described the physical properties of the building as an extension of the openBIM IFC data model [565], [566], through an RDF store and SPARQL endpoint. The semantic model also contextualized the historical data stored in a MySQL database by formalizing a shared meaning. The novel analytics include the automated production of rules through simulation based rule generation [567] and their subsequent fuzzification alongside rules from mining on historical metering data. The visualization component utilized an HTML5 based smart GUI to deliver engaging 3D WebGL visuals alongside real-time and historical energy performance monitoring and decision support, by presenting the optimized rules as user-friendly actuation suggestions. The BEMS aimed to promote trust with facility managers (FM) through a negotiation based user-in-the-loop approach. This meant the FM was responsible for actuating the suggested changes, as this was attractive to industrial partners due to liability and legislation concerns around automated actuation.

The developed solution saved an average of circa 30% energy during the testing period. The current work used the project as a means to engage with industrial stakeholders, mature impact scenarios regarding energy in the built environment, engage with the development of advanced semantically-enabled applications, and analyse the use and performance of the semantic artefacts in the solution.

#### 4.2.1.2 SYSTEM ARCHITECTURE

The project focused on an enhanced BEMS and delivering proof of concept at the selected pilot site. The key components of the proposed system's service-oriented architecture were; the RDF store, SPARQL mapper and knowledge base which constitute the semantic middleware, the data mining engine, rule engine, and fuzzy real time reasoner, which constitute the system's analytics components, and the system's smart GUI, as shown in Figure 26, before a pilot site validation is presented.

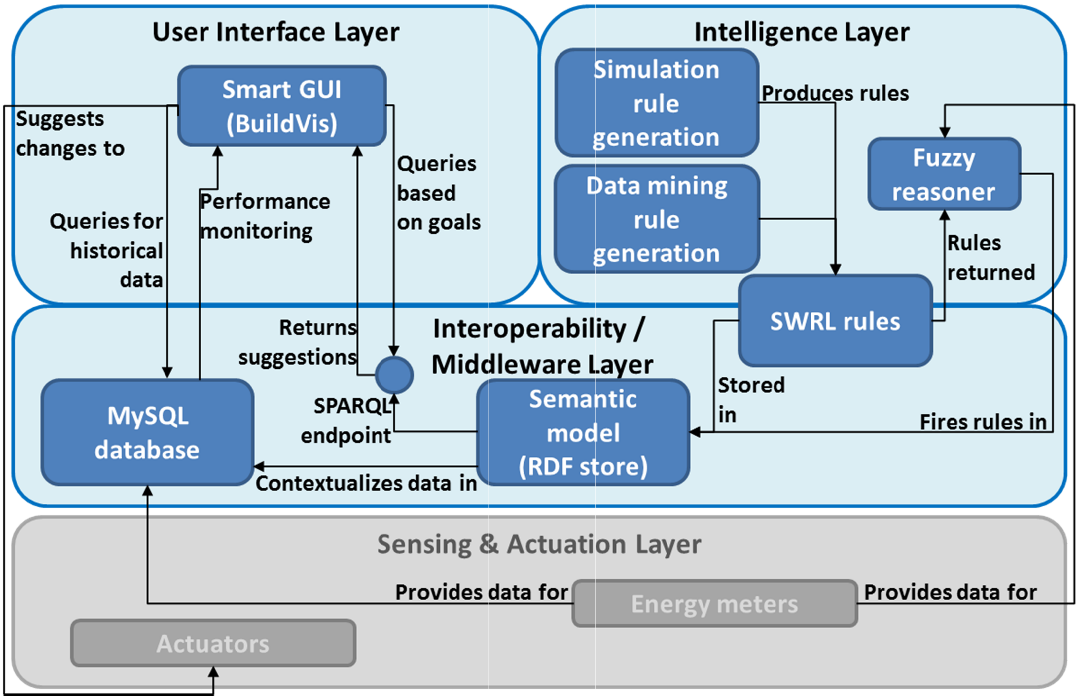


**Figure 27: The architecture of the KnoHolEM BEMS**

The components of the BEMS included a fuzzy logic real time controller (RTC), a graphical user interface (GUI), a mapper, and building-specific systems. Regular, automatic revisions of the rules stored in the knowledge based allowed the system to be dynamic.

Empirical rules were produced through mining of historical sensors data. This aimed to identify correlations between variables, user behaviours, and energy consumption. These were then formalised as SWRL rules in each building's knowledge base. Reasoning on the rules generated new knowledge, such as prediction of the energy consumption of certain user activities, as well as anomaly detection. To consider different aggregation levels, energy consumptions were collected using energy meters at various levels.

The system's intelligence was predominantly derived from its simulation-based optimised rules. This involved pre-processing to produce optimization scenarios and simulation data, and to identify sensitive variables, then training an ANN based on this data. This ANN was then used as the cost function in a GA optimization to output actionable rules, which were evaluated for efficacy.

A public residential care home in The Netherlands, named ‘the Forum’, was the primary pilot site. A thermal simulation model of the building was created in DesignBuilder, as shown in Figure 27 alongside a floorplan highlighting the main zone under consideration.

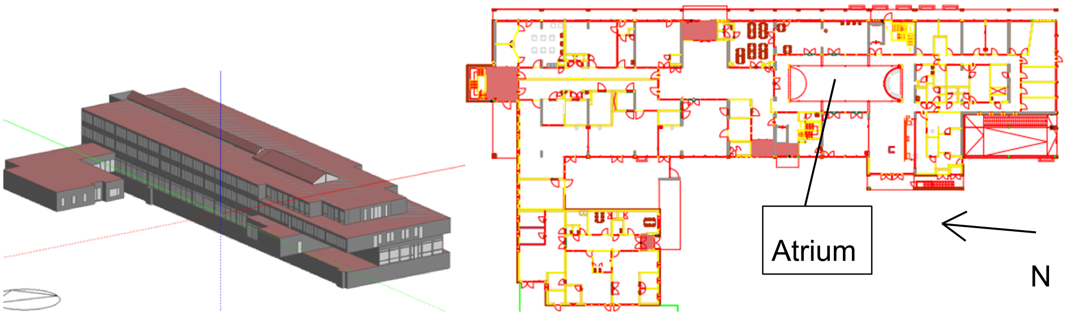


**Figure 28: Energy model of demonstration building and floor plan**

EnergyPlus was then used to produce simulated data across the permutations of the scenario’s independent variables. PCA and MRA were used to reduce the simulation model’s 954 reported variables. The ideal reduction was determined by PCA, and then MRA was used to rank the variables’ sensitivity according to the scenario’s objectives. The identified variables were then mapped with the existing sensors. This data was used to train the ANN model which served as a surrogate model in the cost function of the GA optimisation. As well as the 10 variables identified previously, the ANN included 4 actuator states, and time information, to predict the zone’s PMV and energy consumption. Following several experiments, it was identified that the use of the Levenberg-Marquardt learning algorithm, 1 hidden layer of 30 process elements, and the ‘logsig tansig’ transfer function was ideal. The expected error level of 0.0001 was then achieved after 70 epochs.

To generate the rules, GA optimization was used with an ANN cost function. An ANN model was developed in MATLAB to replace the simulation model due to its speed as a prediction engine, and as the decision space was clearly defined. The building’s actuator states and sensor data were used in the chromosome string. The termination condition of the GA loop was based on input from the FM; the rules generated aimed to reduce the energy consumption by 5, 10, 20 or 30 percent.

Suggestions were produced which identified the actuations required to produce the desired change in building performance.

A fuzzy reasoner module was responsible for running the SWRL rules identified by the former process. This communicated with the GUI through the mapper module, and the knowledge base through the Java Expert System Shell (JESS). The fuzzy reasoner consisted of a fuzzification module, SWRL bridge, rule engine, defuzzification module, and a rule matching module. This reasoner was used when the FM requested decision support. Firstly, the sensor data was compared with the antecedent parts of the rules, then it was fuzzified using triangular membership functions with pre-defined ranges. These fuzzy variables were then processed by the inference engine, which uses SWRL rules from the knowledge base. Defuzzifying the rules then determined the suggested actuation for any given input set.

The project's BEMS interface enabled the FM to monitor a building and utilise the decision support capabilities of the solution, in native web browser languages. Figure 28 shows the WebGL view of the building's zones and Figure 29 shows the Energy Monitoring and Actuation Suggestion window.

The 2D CAD plans of the building were converted to RDF data, and used to make the 3D visualization. As well as showing an extruded floor plan of the building, each zone is described in the knowledge base by its geometric properties, function (kitchen, atrium etc.), ID, and its connected sensors, and these are all displayed after clicking the zone, which triggers a query of the knowledge base.

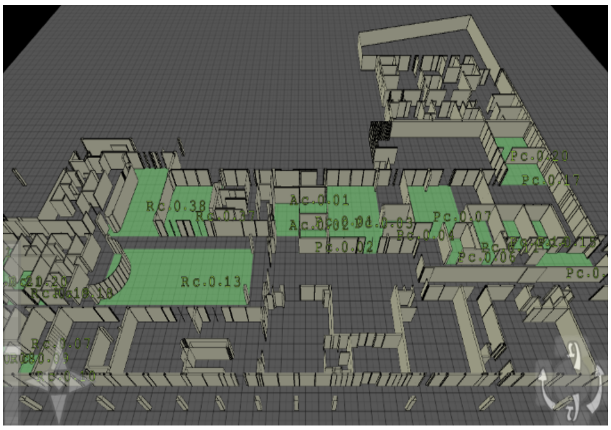


**Figure 29: The WebGL view of the building's zones**

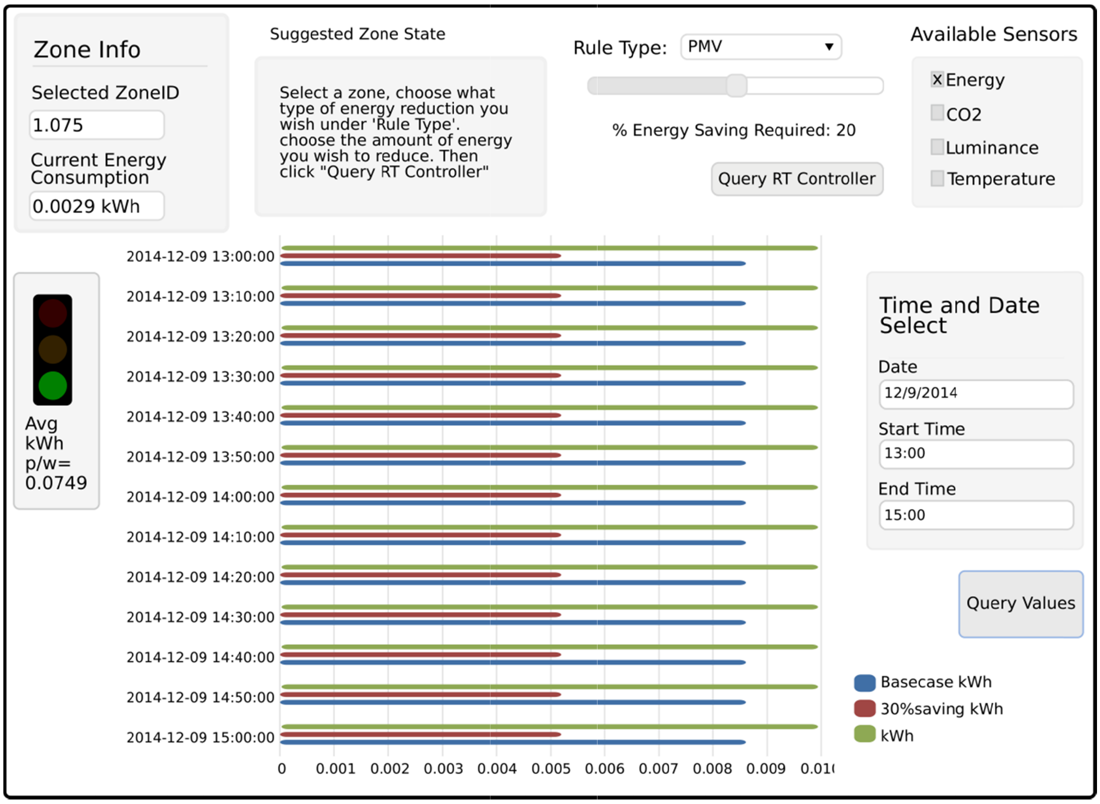


**Figure 30: BEMS zone monitoring interface**

After choosing a zone and a sensor type (or multiple types), the energy monitoring interface shown in Figure 28 allows the FM to view the current and historic performance of the zone. This is achieved through a histogram of sensed data values and a traffic light graphic which indicates the acceptability of the current performance, relative to its mean value. The historical sensor data was retrieved from the SQL database using a combination of AJAX and PHP server-side scripting.

#### 4.2.1.3 SEMANTIC TECHNOLOGIES

The ontological knowledge base was developed by project partners to contain all domain specific information for the KnoHolEM project [568], through building specific instances of a generic domain ontology. It integrated heterogeneous data sources required by the system, and also provides some of the intelligence capabilities through reasoning on the rules and structures contained in the knowledge base. The Web Ontology Language (OWL) was used to represent the knowledge base. The knowledge domain model consists of classes representing building physical elements that are observed and analysed in energy management activities, and building controls consisting of sensors, controllers, alarm, etc., which act as observer and controller of physical building elements. The knowledge model also represents the human actors and their behaviours that can affect the states of building physical elements. This resulted in 145 asserted classes, 43 object property slots and 43 data property slots; the key physical and sensory classes and relationships are shown in Figure 30.

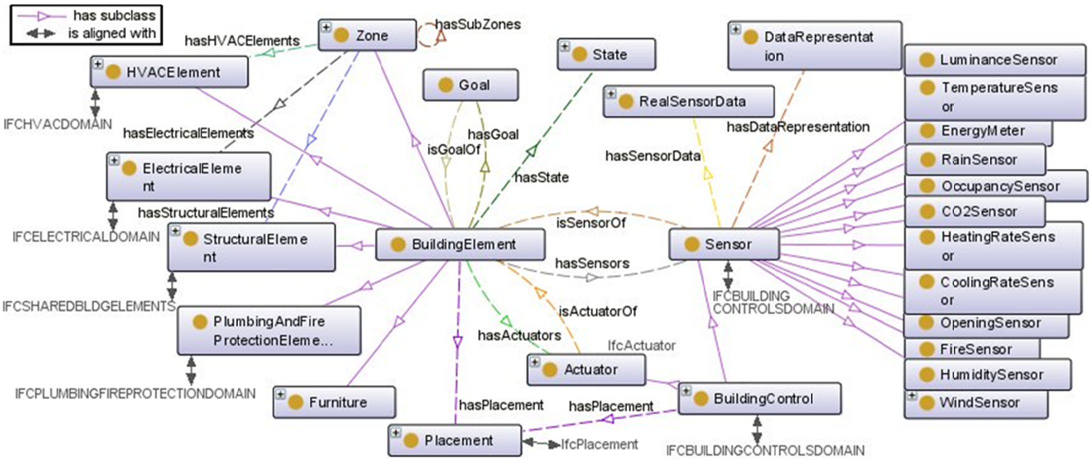


**Figure 31: Main concepts, relationships, and IFC mappings in the BEMS domain ontology**

The ontology is aligned with ifcOWL, as also shown in Fig. 2. For example the IFC entity IfcWindow is mapped to the Window class. The other main IFC concepts which were reused were the physical building elements and geometries such as doors, walls and openings, and the key extensions included descriptions of the zones, sensors, states, people and behaviours in the domain. In total the domain ontology asserted 44 mappings to IFC concepts.

The knowledge base was stored on a Fuseki triple store. To enable visualization of the building floor plan, an existing 2D DWG file is parsed and converted directly into RDF and stored on the Fuseki server. The data extraction tool OntoCAD was used to identify zones in the building and add additional information such as sensor types and locations.

#### 4.2.1.4 OUTPUTS

To evaluate the performance of the developed solution, the system's intelligence was tested in the EnergyPlus simulation environment and the full system was deployed in a real pilot building. The pilot building was a public residential care home in The Netherlands, and the main target zone was the building's 3-storey atrium zone; the main energy consuming space of the building. As the building is primarily an elderly care home, maintaining thermal comfort was critical whilst attempting to reduce the building's energy consumption.

Following a single day simulation, the total energy consumption for the zone was changed from 258 kWh to 201 kWh, and the optimized rules maintained an absolute predicted mean vote (PMV) of less than 1, which was deemed an acceptable level of occupant comfort. The generated PMV and energy consumption profiles from these experiments are shown in Figure 31.

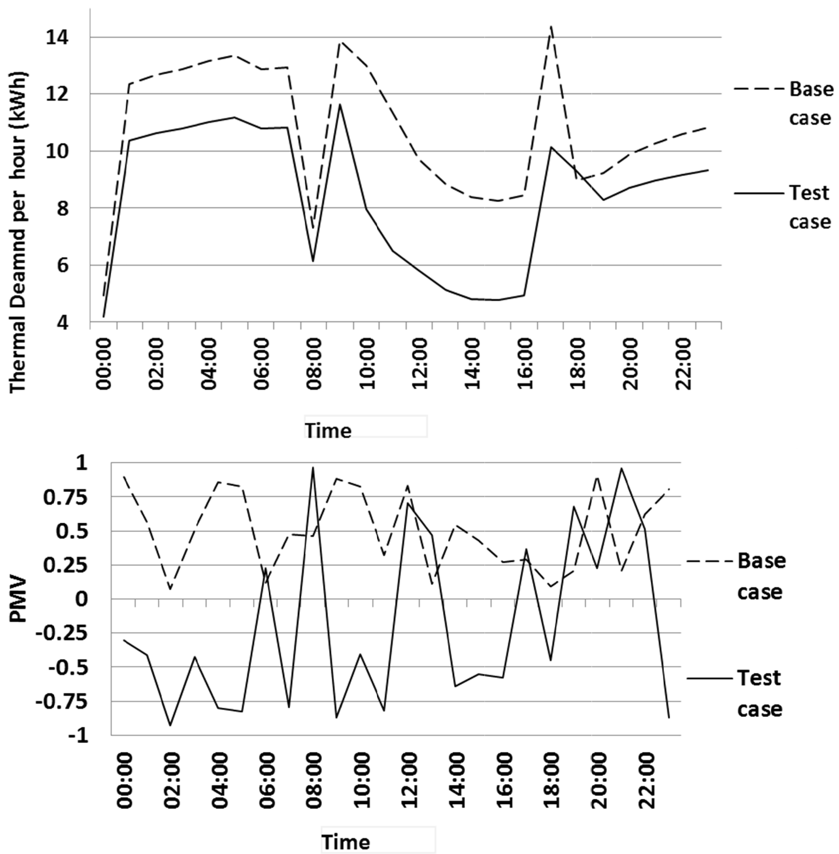


**Figure 32: Simulated effect on single day energy consumption and thermal comfort**

Following preliminary success, the simulation was extended to a two month period. The thermal energy consumption for the simulated building was then found as 11400 kWh using the optimized rules for the months of October and November, compared to total consumption without the rules of 14600 kWh. This represents a 22% reduction, and the relevant energy consumption profiles are shown in Figure 32, where again, the absolute PMV was consistently less than 1.

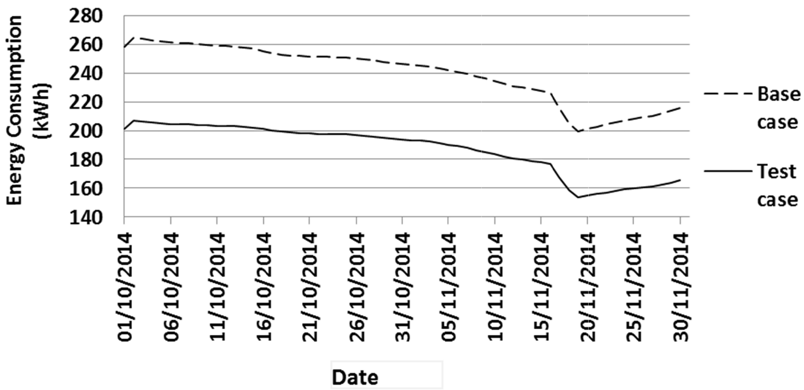


**Figure 33: Simulated effect on daily average energy consumption over two month period**

The full retrofit BEMS solution was then deployed in the pilot building, initially for a single day and subsequently for an extended period from 1 October 2014 to 20 January 2015. In each of these tests the FM utilized the system's decision support to receive suggested actions for energy saving, and after negotiating the severity of these, actioned them through local control systems. Based on the single day experiment, the daily energy consumption was changed from 77kWh to 58 kWh by implementing the proposed system, which represents a 24% reduction, as illustrated in Figure 33.

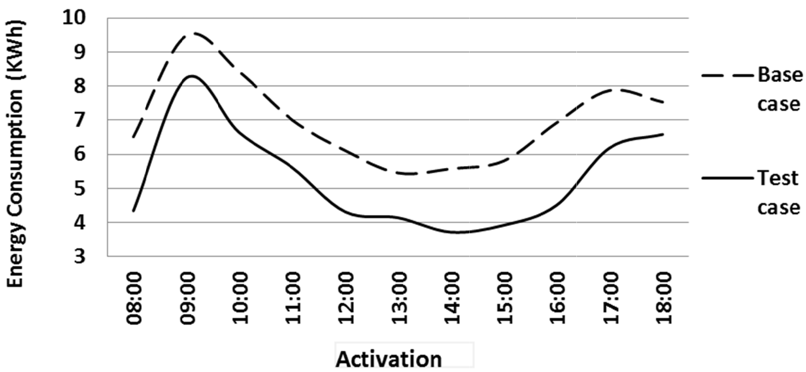


**Figure 34: Effect on pilot site energy consumption over a day**

Further, the total energy consumption during the experiment period in pilot zone was 5580 kWh after implementing the solution, compared to 7510 kWh beforehand, when adjusted for degree day temperature correction. These results are presented in Figure 34 and represent an average winter energy saving of circa 26%. Finally, the FM was satisfied with the thermal comfort achieved and no negative feedback was received from the occupants.

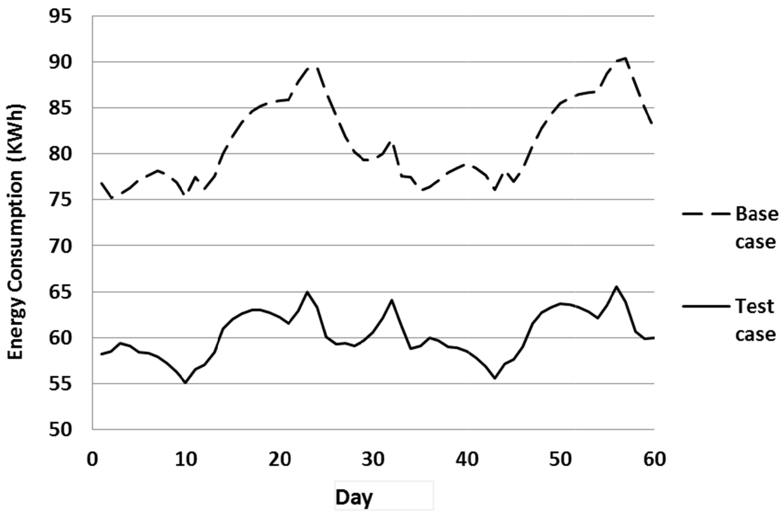


**Figure 35: Pilot building energy consumption over two months**

Following the deployment of the tool at the selected pilot site, the usability of the tool by facility managers was evaluated, and showed that the tool was generally well received, and was sufficient as the output of a proof-of-concept research project.

### 4.2.2 SMART PROSUMER GRIDS

#### 4.2.2.1 OVERVIEW & PROJECT DESCRIPTION

The work on energy management at the building level was extended to consider the optimal management of buildings within their wider power network, in the emerging energy context of prosumers and microgrids. This case study was conducted in the scope of the EC FP7 collaborative research project entitled ‘Multi-Agent Systems and Secured coupling of Telecom and Energy gRIds for Next Generation smartgrid services’ (MAS2TERING).

The project aimed to develop a MAS which was capable of optimising both domestic demand and energy generation and storage vectors through the emergent trading of the virtual commodity of flexibility, thereby creating and testing new markets and business models whilst supporting efficiency and grid resilience. The described MAS was also coupled with novel prosumer forecasting services, and the entire solution utilised a shared ontology and data model to facilitate communication.

The current work utilised this case study as an opportunity to develop and test a semantically enabled solution with a different domain and usage pattern to the previous case study. This offered breadth of analysis through the different application domain, and greater depth of analysis through the actual development, maintenance, and testing, of the ontology.

#### 4.2.2.2 SYSTEM ARCHITECTURE

The MAS2TERING project aimed to facilitate a new market within smart grids through the enablement of flexibility trading as a virtual commodity, within a framework which optimised the system's emergent properties. The project recognises the emerging challenge of DERs as an opportunity to exploit the properties of multi-agent systems, and also the properties of holonic systems. The project developed a MAS instantiation of the USEF framework [569], and extended this through predictive web services and the concept of holonic systems. To support interoperability, an ontological approach was adopted, which used both Java and OWL formalisms of the same ontology. This was required due to the MAS being developed in JADE, and hence requiring the use of a JavaBeans content language ontology, but the wider solution benefitting from the use of a semantic web approach based on an OWL ontology. This approach is illustrated in Figure 35, which shows two agents communicating through the JADE ontology, as well as the home energy management system querying the ontology web service, which would be required to engage with the other web services such as in the use case of predicting the energy generation of the home's micro turbines.

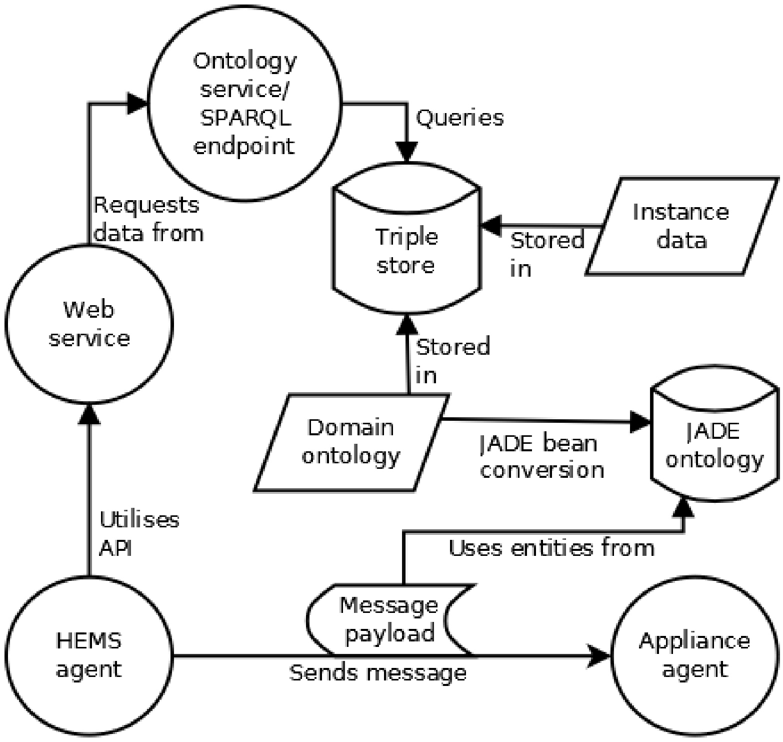


**Figure 36: Illustration of the JADE-OWL ontological approach in MAS2TERING**

The system's primary components are the MAS, predictive web service, and knowledge management. The MAS is primarily based on the USEF framework, and includes agent types for each of the main roles in the framework: device, home, aggregator, and distribution system operator (DSO). Device agents exhibit little agency, and primarily act as abstractions of physical appliances, generation units, and storage devices. They interact with the physical devices, and with users regarding specific physical devices. Home agents then coordinate the devices in the home, based on the flexibility offered by the inhabitants, and based on the energy and flexibility markets offered by the grid and aggregator agents. The aggregator agents then interface between homes and the DSO, by identifying congestion points, setting local flexibility prices, and trading flexibility amongst their profile of homes and with other aggregator agents to improve the net demand profile of the overall area. The DSO agent then interfaces between the aggregators and the external utility at the pilot site, by setting energy and flexibility prices and trading with the aggregators. The predictive web service utilises an ANN to accomplish a data driven prediction of the energy consumption of the houses, and the production of the DERs.

The project investigated the described system across 3 phases; firstly at the building level, by only optimising within individual homes based on fixed pricing schedules. Then the project investigated local energy and flexibility trading up to the aggregator level, to explore the system's behaviour in grid-isolated mode, and finally the full system was explored to evaluate the solution in the business-as-usual case in grid-connected mode.

#### 4.2.2.3 ONTOLOGY DESCRIPTION

The ontology developed aimed to formalise a description of prosumer-enabled smart grids in OWL. This domain can be considered as the emerging energy landscape described in the background section, typified by a high penetration of end users who i) exhibit flexible loads, ii) produce energy, and iii) store energy, as well as being typified by a high penetration of renewables, microgrids, virtual power plants, plug-in electric vehicles, polygeneration systems, and district heating networks. The ontology integrated existing standards in the smart grid and demand side management domains, as well as lessons learnt in other case studies.

The data model required in MAS2TERING facilitated the common expression of information exchange between participating entities. Instead of multiple ad-hoc mapping and conversion processes between arbitrary models, participants will either use the common model internally or map their internal model to a common schema. The common ontology will be used for formulating messaging data structures for syntactical and implied semantic compatibility between entities for the support of upper business processes. The common schema borrowed constructs from current standards. Three primary standards were identified for use, with varying relevance across the project's use cases; namely the IEC 61970 standard for modelling the electrical domain, the OpenADR standard for modelling demand response within the smart grid, and the energy@home standard for domestic conceptual modelling. Subsequently, the IFC and SAREF models were also used and aligned with.

Due to the predominance of multi agents in the MAS2TERING platform the primary usage of the data model was to formulate content for those messages, independent of protocols. As such the conceptual modelling resulted in an ontology suitable for use within JADE. This extended JADE's Base Ontology Java class to formalize a vocabulary, as well as descriptions of the concepts, predicates, agent actions and data slots relevant to the domain. In addition schemas formed the ontology metadata attributes of those messages and the schemas were generated from defined custom contexts applied to the base models. Both extending JADE's Base Ontology and defining and registering the schemas were achieved by means of a JADE BeanOntology, given that concept, predicate and agent action beans were produced first.

Following a thorough analysis of the standards, these were then federated manually into ontological representations using OWL constructs. Subsets of these ontologies were then aligned and extended to fully model the MAS2TERING domain. The main areas of extension regarded optimisation, device types and descriptions, demand response and load control. The results of the use case based elicitation are presented in the following sections followed by the resultant ontology.

#### 4.2.2.3.1 USE CASE BASED DESCRIPTION LOGIC ELICITATION

Based on the methodology's use case driven approach to elicit a lightweight ontology aligned with existing standards, each use case was considered in turn, with concepts, relationships and properties being elicited to satisfy exchange requirements. These were then compared to existing standards to determine potential alignments before formalising the MAS2TERING ontology.

Regarding the use case of smart home energy management, the energy@home data model was reused and extended. The use case of aggregating several smart homes aimed to utilise the flexibility offered by consumers within a new market. Aggregator agents receive a prosumer-plan (P-plan) from each home, buy the required flexibility, and then trade flexibility with each other, and relay these requests to the smart homes. This use case required the modelling of multi-home concepts, and flexibility concepts. This reused significant parts of the openADR and CIM standards.

The final use case aimed to reduce the number and impact of congestion points within a low voltage grid through the introduction of DSO and facilitatory agents. These agents exchanged knowledge and traded flexibility with aggregators in order to improve the efficiency and resilience of the grid. This required the modelling of connections and congestion points, as well as DSOs and further modelling of flexibility. As the concept of flexibility was central to the approach, a clear understanding of its definition and a thorough formalisation of its nature was required, which is specified in the following section.

#### 4.2.2.3.2 DOMAIN PERSPECTIVE OF ENERGY FLEXIBILITY

Load flexibility was here defined as *'the quantity of peak load which is reduced through optional deferment and/or curtailment of consumer demand, expressed as a unit of energy; this is typically sold by energy consumers for the betterment of grid management'*. Deferment is '*the shifting of a load to a time more favourable to the network operator, where the amount of flexibility is equal to the amount of energy shifted'*. In this way the extent of the shift is independent to the flexibility, as the consumer sets a deadline for the task completion. This is represented in Figure 36 below, when Qtot is the total energy consumption of the task, Qf is the flexibility utilised, t0 is the earliest start time of the task, t1 is the task completion deadline, and Tmin is the minimum amount of time the task requires to be completed.

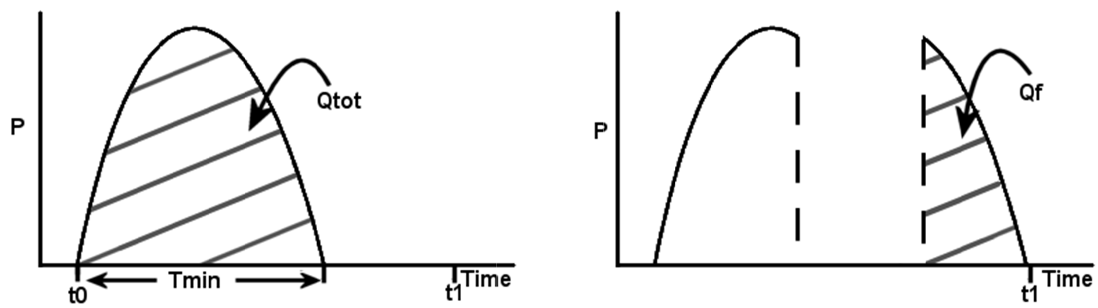


**Figure 37: Ontological perspective of load deferment. Left - desired load, right - deferred load**

Curtailment of load is then *'the supply of a quantity of energy over time which is less than the desired quantity'*. The flexibility is then the difference between the desired quantity and the supplied quantity, again expressed as an amount of energy. This is shown in Figure 37, where t0 is the earliest start time of the task, t1 is the non-negotiable deadline of the task, Qf is the amount of flexibility utilised, and Pmin is the minimum amount of energy to be supplied (such as when a heating device must meet a minimum room temperature).

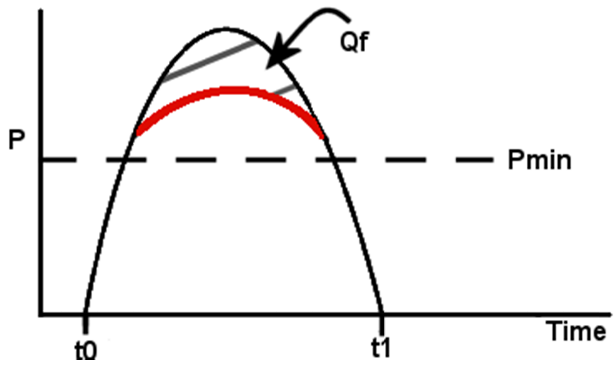


**Figure 38: Domain perspective of load curtailment. Black profile - desired load, red line - curtailed load**

Devices were categorised according to their likely flexibilities, as shown in Table 22 below.

**Table 22: Classifications of likely flexibilities of devices.**

| | Non-interruptible | Interruptible | Variable Profile |
|---|---|---|---|
| Curtailable | N/A | N/A | Electric heater |
| Deferrable | Washing machine | Dishwasher | PEV, electric oven, tumble dryer, kettle |
| Fixed | Freezer, fridge, lights | N/A | N/A |

### 4.2.2.3.3 CANDIDATE DOMAIN ONTOLOGY

Following the elicitation of domain knowledge, this was then formalised in description logic using basic OWL constructs so as to produce a candidate ontology, through the use of Protégé software. The ontology is presented first below using OWL constructs, and its formalisation in the JADE format utilised in the MAS is subsequently mentioned. Figure 38 - Figure 40 and Table 23 present the main concepts, relationships and data properties formalised. The energy scheduling model view definition (MVD) formalises energy scheduling classes, as specified by the energy@home data model [471], which expresses that a power profile is made up of an ordered set of modes, which are then composed of energy phases.

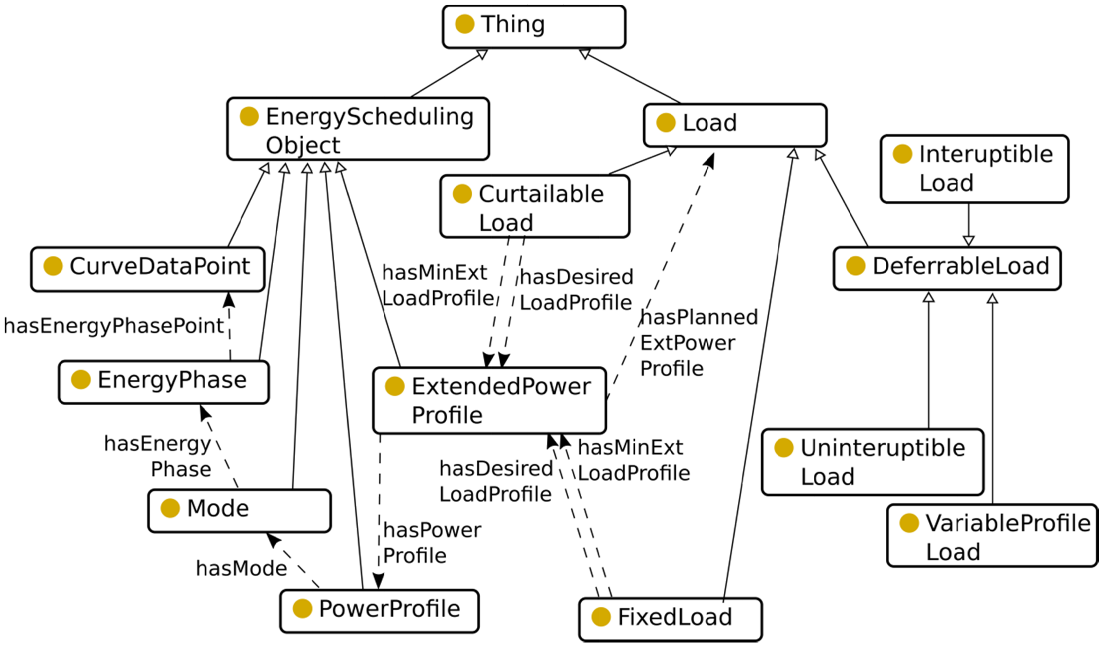


**Figure 39: MAS2TERING OWL energy scheduling MVD**

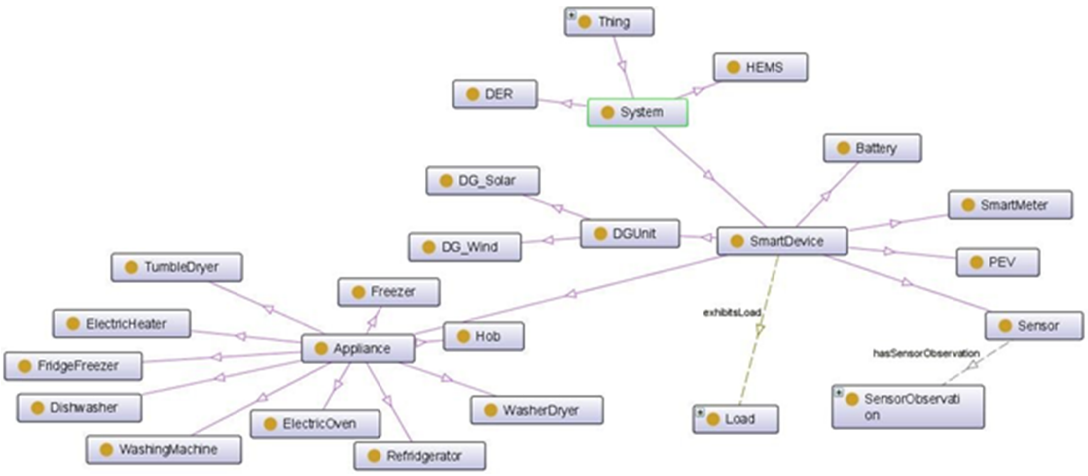


**Figure 40: MAS2TERING OWL device MVD**

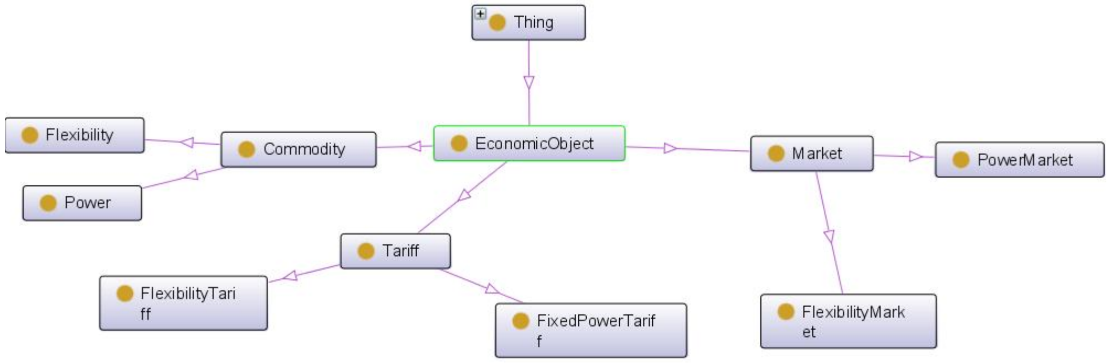


**Figure 41: MAS2TERING OWL economic MVD**

**Table 23: MAS2TERING MVD data properties**

| Data Property | Functional | Domain | Range |
|---|---|---|---|
| **hasSupplyUnitPrice** | ✓ | FixedPowerTariff | float |
| **hasMaxOverloadPause** | ✓ | EnergyPhase | float |
| **hasTotalEnergyDemand** | ✓ | DeferrableLoad | float |
| **hasDemandUnitPrice** | ✓ | FixedPowerTariff | float |
| **hasActivePower** | ✓ | CurveDataPoint | float |
| **hasDuration** | ✓ | EnergyPhase, ExtendedPowerProfile, Mode, PowerProfile | float |
| **hasHeatingSetPoint** | ✓ | ElectricHeater | float |
| **hasTimestamp** | ✓ | SensorObservation, CurveDataPoint | string |
| **hasMaxDischargeRate** | ✓ | Battery | float |
| **hasMaxChargeRate** | ✓ | Battery | float |
| **hasSequenceOrder** | ✓ | EnergyPhase, PowerProfile | int |
| **hasMaxDelay** | ✓ | EnergyPhase | float |
| **hasPeakPower** | ✓ | EnergyPhase | float |
| **penaltyParameter** | ✓ | CurtailableLoad | float |
| **applianceID** | ✓ | Appliance | string |
| **hasExecutionDeadline** | ✓ | DeferrableLoad | string |
| **hasMinDuration** | ✓ | DeferrableLoad | float |
| **hasBatteryCapacity** | ✓ | Battery | float |
| **hasBatteryInitialCharge** | ✓ | Battery | float |
| **hasMinPowerProfileDelay** | ✓ | PowerProfile | int |
| **hasSensorValue** | ✓ | SensorObservation | float |

### 4.2.2.3.4 ALIGNMENT WITH EXISTING STANDARDS

The ontology was aligned with existing standards due to reusing them, which is introduced in Table 23. This doesn't represent full compliance or alignment with the standards, but it demonstrates broad coherence with their domain perspectives of the existing standards.

**Table 24: Alignment of MAS2TERING terms entities with existing specifications**

| MAS2TERING entity | IEC 61968-9 alignment | CIM alignment | energy@home alignment |
|---|---|---|---|
| Dwelling | UsagePoint | | |
| FixedPowerTariff | Tariff | | |
| HEMS | LoadControlDevice, Head End, PANDevice | EnergyConsumer | |
| IntervalBlock | IntervalBlock | | |
| MeterReading | MeterReading | | MeterReading |
| Reading | Reading | | |
| SmartDevice | EndDevice | | EndDevice |
| SmartMeter | Meter | | |
| Tariff | PricingStructure | | |
| AggregatedDwelling | | LoadGroup, ControlArea | |
| DG_Wind | | WindGeneratingUnit | |
| DGUnit | | GeneratingUnit | |
| DSO | | ControlArea | |
| Dwelling | | ControlArea | |
| hasLimit | | Limit | |
| hasReadingValue | | MeasurementValue | |
| hasSetPoint | | SetPoint | |
| Sensor | | Sensor | |
| SensorObservation | | Measurement | |
| Appliance | | | Appliance |
| CurveDataPoint | | | CurveData |
| EnergyPhase | | | EnergyPhase |
| ExtendedPowerProfile | | | ExtendedPowerProfile |

| | |
|---|---|
| Mode | Mode |
| PowerProfile | Mode |

The outputs of the case study were threefold; firstly, the ontology itself represents an output, as it aligns several seminal models for a valuable purpose, secondly, the integration of the ontology into the knowledge management software and subsequently into the wider ICT solution produced useful knowledge about the role of advanced applications and semantics in the domain. Finally, the case study reused the ontological artefacts produced to further explore the role of reuse and semantics in smart cities, which represents a significant output.

## 4.3 WATER SECTOR INVESTIGATION

### 4.3.1 OVERVIEW & PROJECT DESCRIPTION

This case study tested the thesis' hypothesis in an industry with a less mature smart technology market, by developing a semantically-enabled cloud platform for water knowledge management. This involved the development, validation, and experimentation of a domain ontology and knowledge-based system for the water sector, within the context of an EC FP7 project, entitled 'Water analytics and Intelligent Sensing for Demand Optimised Management' (WISDOM). The case study built on the lessons learnt and artefact developed in the previous case studies in the energy sector.

The present study utilised this project to gather industrial expert domain perspectives, as well as providing real-world use cases whilst building and testing a knowledge-management solution and domain ontology. The work conducted went beyond the remit of WISDOM to test the knowledge-management solution in a standalone knowledge-based system, developed specifically to evaluate the hypothesis.

WISDOM aimed to demonstrate the value of applying ICT to the water domain in operational systems, by installing sensors and using a cloud-edge system to support advanced applications. These applications included intelligent leakage detection, pumping optimisation, and sewer overflow spill prediction, which used AI and optimisation software to support decision makers at pilot sites in Wales, France, and Italy.

The work conducted was a thorough requirements engineering process in close collaboration with experts, before developing a domain ontology and accompanying knowledge-based system and inference engine. The artefacts produced and the evidence gained in engaging with industrial experts was used to evaluate the validity of the hypothesis. The knowledge base can import GIS data and integrate this with dynamic sensor data, social concepts and inference rules.

The core element of this case study; the domain ontology, was engineered following the guidance of the well-established NeOn methodology [570], and represents a formal and shared description of the concepts and relationships in the domain of water management. These concepts are grouped into a Water Catchment Information Model, a Water Semantic Sensor Network Ontology and a Water Value Chain Social Model. The domain ontology and ontology service have undergone a multi-stage validation process. The instantiation of the knowledge bases was accomplished through a semi-automated method of legacy system integration.

### 4.3.2 SYSTEM ARCHITECTURE

The system is conceptually arranged into 4 architectural layers: sensing infrastructure, data acquisition and actuation, core services and business services, as shown in Figure 42. The core services layer contains the system's semantic integration service, optimization and analytics services, event bus and governance module. These core components utilize data communicated from the sensing infrastructure to the event bus via the data acquisition layer, and are delivered to users through the GUIs and edge analytics which form the business service layer. The intended users of the system would primarily be operational staff of the water or waste service provider, although consumers may have access to a restricted set of services. The key innovation considered herein is the use of the core services to integrate analytics across heterogeneous data sources by standardizing data syntax and meaning. However, one of the main applications supported by the semantic web service is now described.

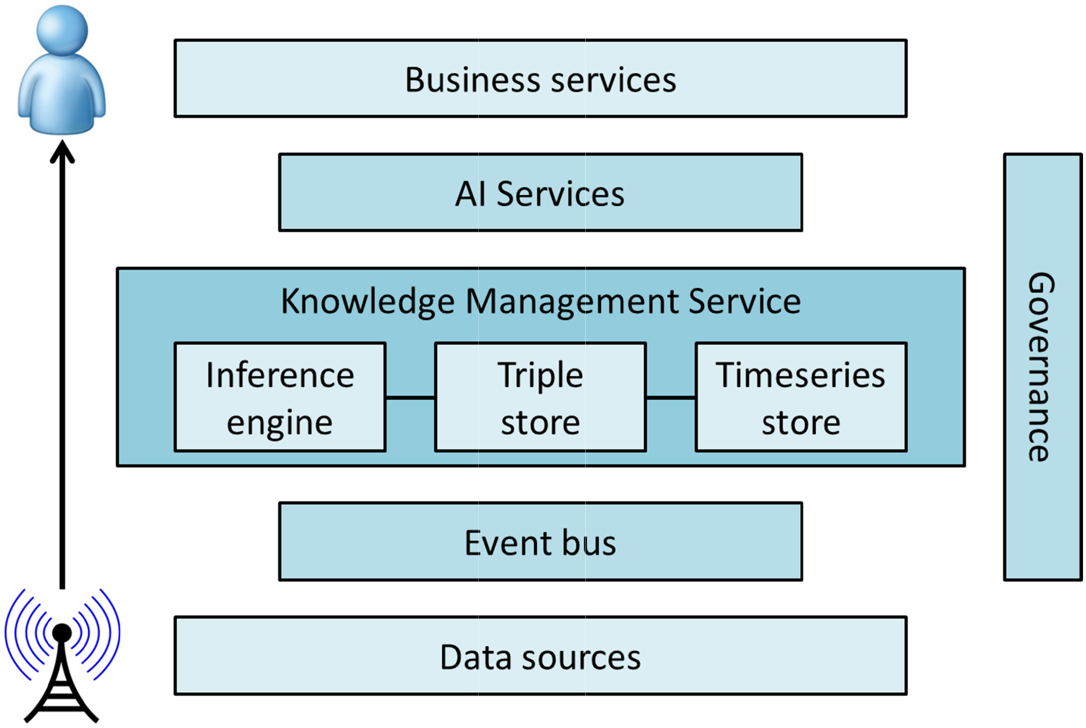


**Figure 42: Architecture of the proposed ICT solution**

## 4.3.3 REQUIREMENTS ENGINEERING

The first stage of the case study was to specify the project's requirements. This section summarizes the outputs of this process of eliciting requirements for the semantic models and the accompanying software solution.

### 4.3.3.1 PRELIMINARY META-REQUIREMENTS

The requirements engineering process began by producing a set of meta-requirements which determined when the requirement specification was acceptable, in terms of their testability and clarity. Example meta-requirements are presented below:

- A requirement must describe clearly what is required of the software solution such that it can facilitate the achievement of each scenario's goals.
- A requirement must describe the delivered software solution, rather than the outcome of the overall project or any of the development processes within the project.
- A requirement must utilise appropriate language such that requirements are testable.

#### 4.3.3.2 INITIAL ONTOLOGY SCOPING

The next output of the requirements engineering process was a set of informal statements about the scope of the ontology. Examples are presented below:

- Should model the entire water network's physical and topological properties
- Should model the actuation and instrumentation layer, and its relationship to water network
- Should be aligned with existing models such as W3C, SSN, WaterML and emerging efforts such as SWIM
- Electricity consumption will be included, but its management is not a focal point
- Managing the internal operation of treatment plants and pumping stations is not a focal point

#### 4.3.3.3 COMPETENCY QUESTIONS

The competency questions were produced iteratively alongside the project's scenarios, by considering the main entities and the properties within each scenario. Questions were then formed which elicited these properties, as illustrated in Table 24.

**Table 25: Example competency question elicitation process**

| **Scenario 1 (Behaviour & Feedback)** | |
|---|---|
| Question | *How much water does person X consume per week, on average?* |
| Property | Average weekly water consumption |
| Entity | Domestic resident |
| **Scenario 1 (Behaviour & Feedback)** | |
| Question | *Which water meter is attached to house X?* |
| Property | Attached water meter |
| Entity | Domicile, domestic water meter |
| **Scenario 2 (Network monitoring)** | |
| Question | *What property does sensor X detect?* |
| Property | Observed property |
| Entity | Sensor |
| **Scenario 11 (Reservoir optimization)** | |

| | |
|---|---|
| Question | *What is the maximum storage volume of service reservoir X?* |
| Property | Max storage volume |
| Entity | Service reservoir |

### 4.3.4 ONTOLOGICAL DOMAIN MODEL

This section presents the semantic domain models which have been produced, split into three conceptual regions; the Water Catchment Information Model (WCIM), Water Semantic Sensor Network Ontology (WSSNO) and the Water Value Chain Social Model (WVCSM). These represent a language used to produce knowledge bases for each of the pilot sites. Firstly the meta-model developed from the W3C SSN ontology and the STS ontology is presented before discussing and showing each model in turn.

#### 4.3.4.1 DOMAIN INDEPENDENT META-MODEL

Following the alignment and extension of the reused ontologies, a domain independent meta-model was produced which described the concepts and relationships in the upper domain of intelligent sensing in a socio-technical network. The choice of a socio-technical system approach was made due to both the autonomy exhibited in the water network's social subsystems and the clear divide between physical, technological concepts and social concepts. This implied that the modelling view should go beyond a systems theory approach and even beyond a system of systems approach to a socio-technical systems approach, and hence reused the work of Koen van Dam [550], the main relevant components of which are shown in Figure 43.

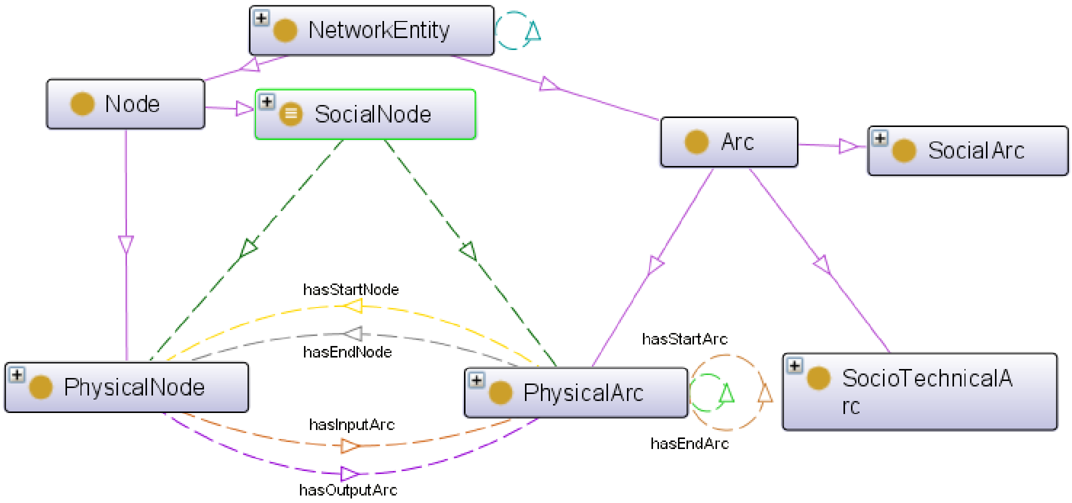


**Figure 43: Main reused concepts and relationships from the STS ontology**

The use of the W3C SSN ontology was a natural choice due to the heavy reliance on intelligent sensing in WISDOM, and so this was merged with the STS ontology through an alignment process. The resultant meta-model was extended slightly with object properties and object property restrictions, and an excerpt of it is shown in Figure 44.

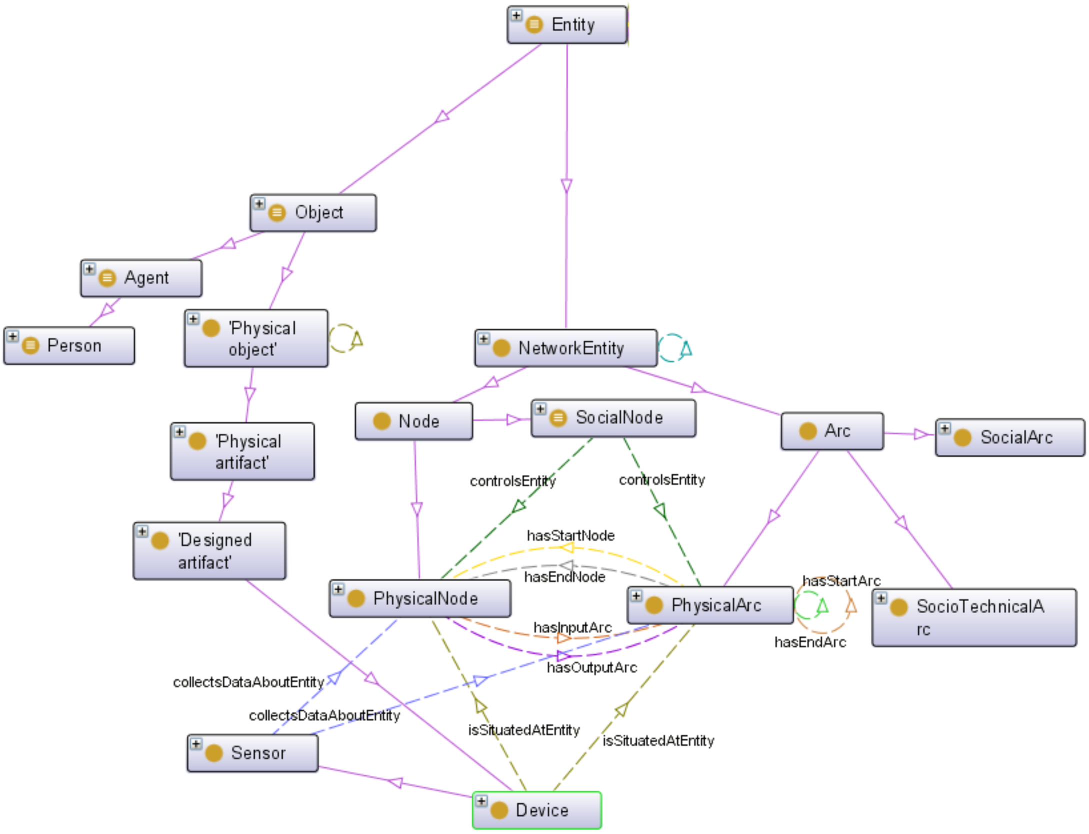


**Figure 44: Excerpt of the meta-model for intelligent sensing in socio-technical systems**

### 4.3.4.2 WATER CATCHMENT INFORMATION MODEL

The Water Catchment Information Model (WCIM) describes the concepts and relationships relevant to the physical infrastructure in the water value chain. This was developed from business process modelling, GIS data schemas, and the preliminary knowledge acquisition. The WCIM model can be conceptually divided into 5 models; supply-side artefacts, water network topology, process model, domestic artefacts, and natural water body descriptions. These components are described through the following sub-sections.

#### 4.3.4.2.1 THE WATER VALUE CHAIN PROCESSES

Figure 45 below shows the process model of the water value chain's key processes. This was used to categorise physical components by purpose. From Figure 45, each process has been modelled as a concept; all technical components *facilitate* one of these processes, and each process has input and output substances (water types or by-products such as sludge, waste solids, or rag & grit).

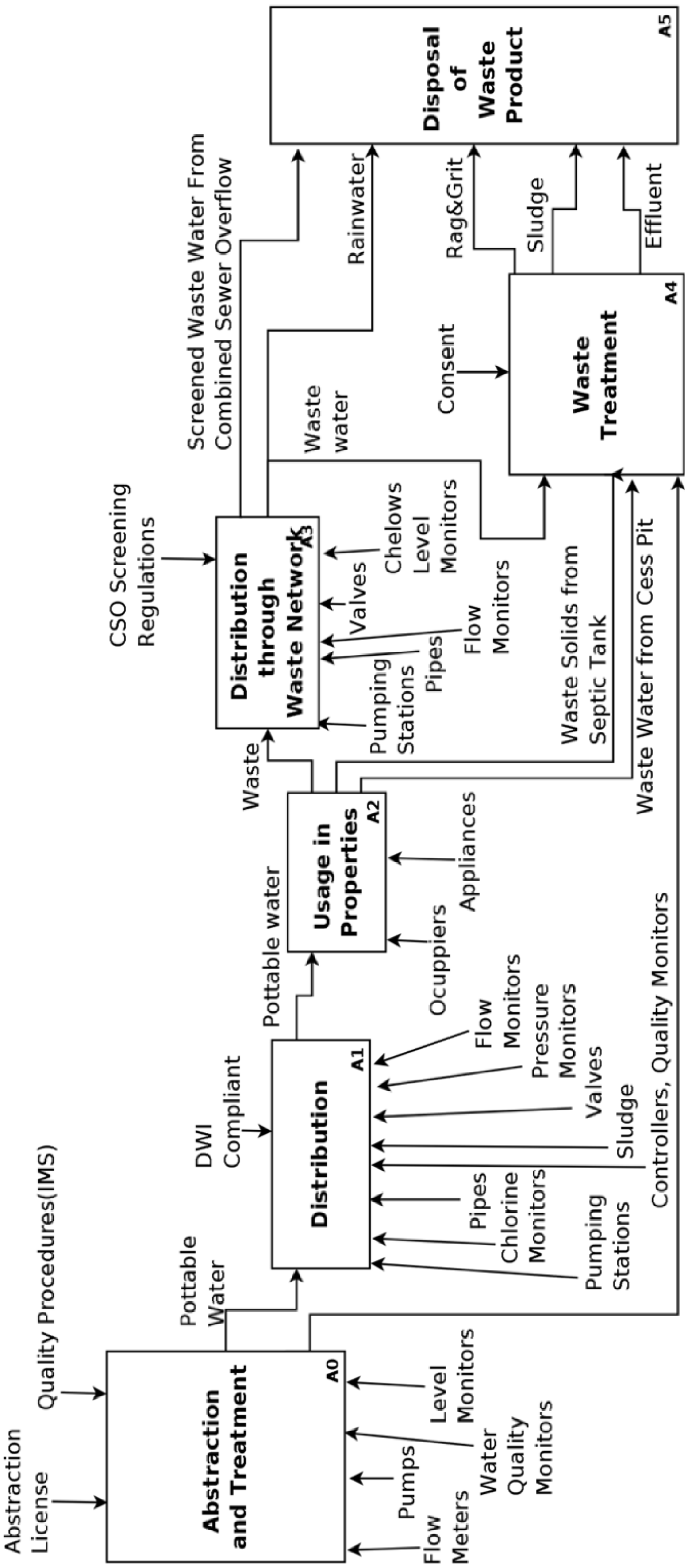


**Figure 45: Processes in the water value chain**

### 4.3.4.2.2 WATER NETWORK MODEL

A water network is modelled as a collection of nodes connected by arcs, where nodes represent utility assets and arcs represent pipes, and each has a taxonomy of object types. Examples of nodes are pumping stations, reservoirs, and buildings. Note that devices at assets such as individual pumps are out of scope. All node classes are subtypes of the generic node presented in Figure 46 below. Each node is also described by its geographic coordinates, elevation, and by its preceding and succeeding arcs. Natural water bodies are modelled separately.

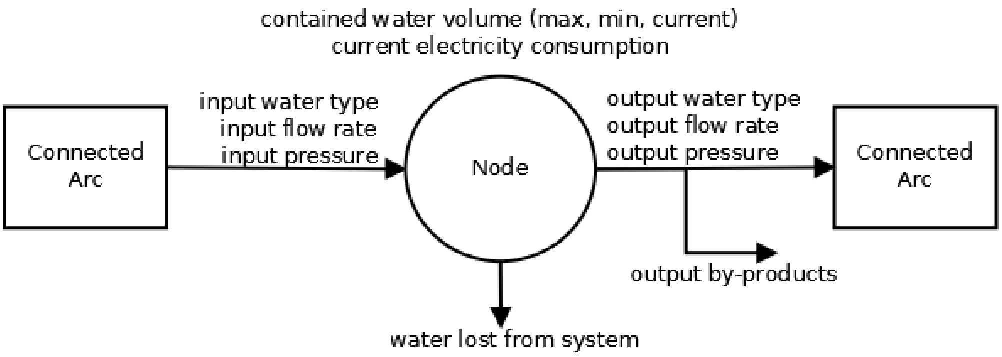


**Figure 46: Description of a generic node and its relevant data**

From this generic node description, subtypes are described by the restrictions they place on aspects of the generic node, shown in Table 25 below. 'Abstraction node' and 'discharge node' refer to the start and end of the man-made water network. These correspond to the abstraction and discharge points, not to natural water bodies. They are instead related to natural water bodies through '*abstractsFrom'* and '*dischargesTo'* relationships.

**Table 26: Subtypes of the generic node**

| Node subtype | Restrictions | Extra properties |
|---|---|---|
| Abstraction node | Output type must be 'raw water'.<br>Contained water volume is assumed to be infinite.<br>Has no input. | Max output |
| Discharge node | Max contained water volume is assumed to be infinite.<br>Has no output. | |

| | | |
|---|---|---|
| Pump node | Output water pressure must be higher than input.<br>Input flow must equal output flow.<br>Input type must equal output type.<br>Contained water volume is assumed to equal zero. | Currently active |
| Storage node | Input type must equal output type.<br>Max contained water volume must not equal zero. | Max stage height, current stage height, min stage height |
| Consumption node | Output type must be ‘foul water’. | |
| Treatment node | Output type must not equal input type. | Max output |
| Valve node | Contained water volume is assumed to equal zero.<br>Input type must equal output type. | Current degree of openness |

Water and waste pipes are modelled as arcs. Different sizes and types of pipe are subtypes of the generic pipe shown in Figure 47.

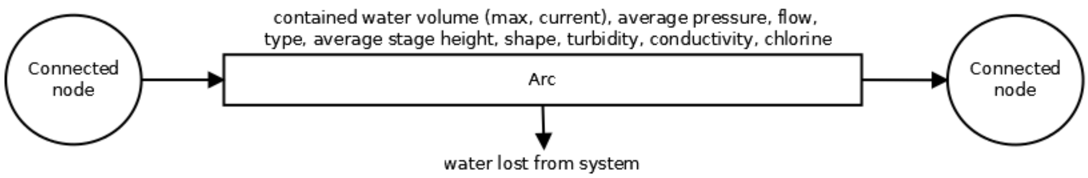


**Figure 47: Description of a generic pipe and its relevant data**

Each fitting between pipes is considered as a node, but whole stretches of pipe between assets can be modelled as a single pipe if suitable for the target application, with sections then modelled as sub-arcs. Pipes are classified as either clean, raw, or waste, and are classified further as shown in Figure 92. Pipes are also described by their material, IPID, diameter (absolute and nominal), length, starting and finishing coordinates and elevations, and the preceding and succeeding nodes or pipes. Also, pipe shapes are modelled as object properties.

A water network consists of only nodes and arcs, where each arc can be connected between only two nodes, but nodes can have any number of input and output arcs. Natural water bodies are related through abstraction and discharge nodes, and described through data properties. A fictitious water network using this approach is shown in Figure 48.

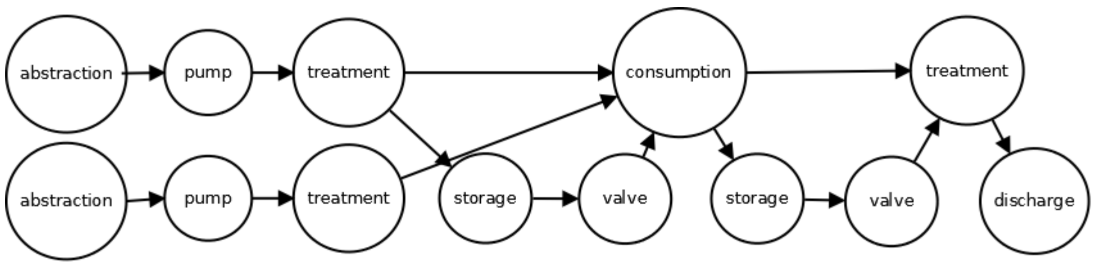


**Figure 48: Simple water value chain showing connection of nodes (labelled) and arcs (arrows)**

The main upper classes and relationships of the WCIM are illustrated in Figure 49.

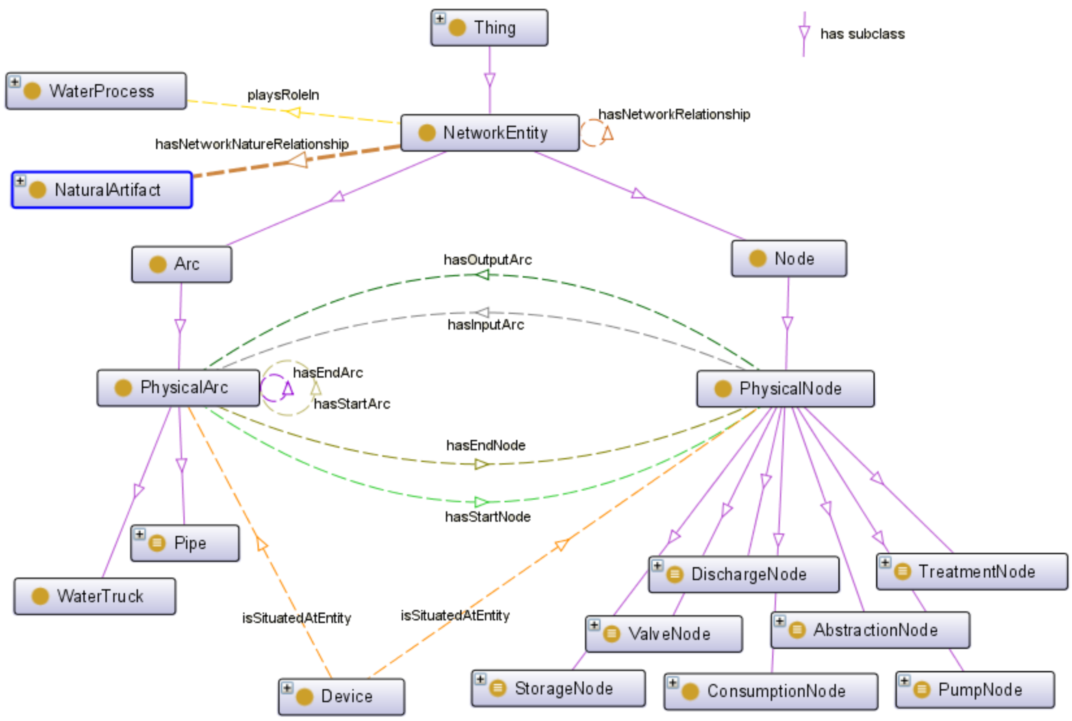


**Figure 49: Main WCIM level classes and relationships**

### 4.3.4.2.3 SUPPLY-SIDE ARTEFACTS

Table 26 details the main artefacts and node types within each process of the water value chain, and is illustrated in Appendix A.

**Table 27: Artefacts in each water value chain process**

| Process | Artefact | Node type | Sub-types |
|---|---|---|---|
| **Abstraction** | Raw pumping station | Pump node | |
| | Raw water pipe | N/A (arc) | |
| | Raw water vehicle | N/A (arc) | |
| **Clean treatment** | Clean water treatment plant | Treatment node | |
| **Clean distribution** | Clean main | N/A (arc) | Communication pipe, Distribution pipe, Trunk main |
| | Water vehicle | N/A (arc) | |
| | Clean storage device | Storage node | Service reservoir, water tower |
| | Clean pumping station | Pump node | |
| | Clean valve | Valve node | Boundary valve, control valve, pressure regulating valve |
| **Consumption** | Hydrant | Consumption node | |
| | Boundary box | Valve node | |
| | Building | Consumption node | Commercial building, Industrial building, Domestic building, Other building |
| | Domicile | Consumption node | |
| **Waste distribution** | Waste main | N/A (arc) | Combined waste main, connecting |

| | | | |
|---|---|---|---|
| | | | sewer, foul pipe, lateral drain, surface water pipe |
| | Waste truck | N/A (arc) | |
| | Chamber | Storage node | Combined chamber, foul chamber, storm water chamber, surface water chamber, treated effluent chamber |
| | Waste pumping station | Pump node | |
| **Waste treatment** | Waste treatment works | Treatment node | combined treatment works, foul treatment works, surface treatment works |
| **Discharge** | Discharge pipe | N/A (arc) | Storm overflow pipe, emergency overflow pipe, treated effluent pipe |

#### 4.3.4.2.4 DOMESTIC ARTEFACTS

A key goal of the WISDOM project was to integrate knowledge across supply and demand, including smart meter data. Concepts were therefore modelled around domestic water consumption devices, water-saving devices and greywater devices. As water flows between devices are not, often, actively managed in domestic properties, domestic pipes were not represented. Table 27 describes relevant artefacts, and Figure 93 presents the OWL class hierarchy of this.

**Table 28: Artefacts relevant to domestic water consumers**

| Process | Artefact | Sub-types | Extra properties |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Water usage | Water using appliance | tap, bath, shower, toilet, washing machine, dishwasher, irrigation system, boiler, other | Average water consumption per use, average energy consumption per use |
| Rainwater usage | Rainwater harvesting device | | |
| | Rainwater purification device | | |
| Waste storage | Domestic waste storage tank | Septic tank, Cess pit | |
| Greywater usage | Greywater harvesting device | Reuse device, harvesting device | |
| Water metering | Domestic water meter | | |

#### 4.3.4.3 SENSOR ONTOLOGY

The sensor ontology described the concepts and relationships relevant to IoT and web-enabled telemetry systems, by adapting the Semantic Sensor Network (SSN) ontology. The SSN ontology models sensors as devices which implement a ‘sensing process’ through a ‘measurement capability’. The ‘measurement capability’ is subject to various conditions and properties regarding operating conditions and accuracy. The sensing process receives a stimulus from the sensed event and outputs an information object describing some property of a feature of interest. This is illustrated in Figure 50, and described further in Table 28.

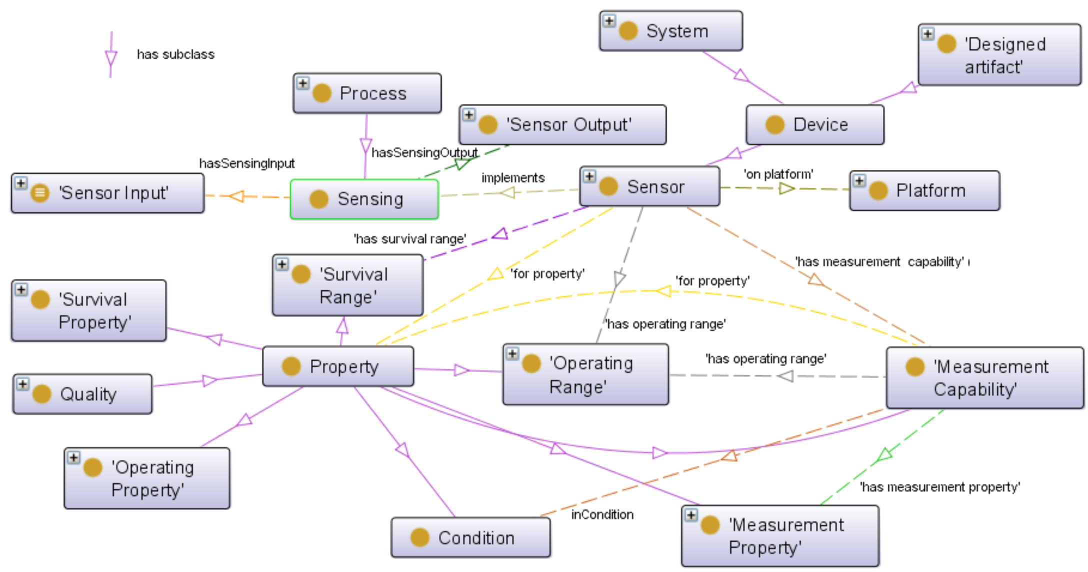


**Figure 50: OWL implementation of the SSN ontology**

**Table 29: Descriptions of main ambiguous SSN classes**

| Class | Description |
|---|---|
| **Stimulus** | The event which is the input of the sensing process |
| **Operating range** | The environmental conditions and characteristics of a system/sensor's normal operating environment. |
| **Survival range** | The conditions a sensor can be exposed to without damage. |
| **Measurement property** | A characteristic of a sensor's output. e.g. Accuracy, detection limit, drift, frequency, latency, measurement range, precision, resolution, response time, selectivity, sensitivity. |
| **Observed property** | The quality of the physical phenomenon sensed which the sensor observes. |
| **Sensing** | A process implemented by a sensor which produces an information object describing the value of a property of a phenomenon. |
| **Measurement capability** | Collects together measurement properties (accuracy, range, precision, etc.) and the environmental conditions in which those properties hold, representing a specification of a sensor's capability in those conditions. |
| **Condition** | Used to specify ranges for qualities that act as conditions on a system/sensor's operation. |

The SSN ontology does not include any measurement concepts (units, etc.), nor water specific taxonomies, so has been extended to include these. An excerpt of this extension is shown in Figure 51, where all arrows indicate a ‘*hasSubclass’* relationship. Water sensors also have additional data properties, including ‘*ProductNumber’*, ‘*Manufacturer’* and ‘*AttachmentType’* etc.

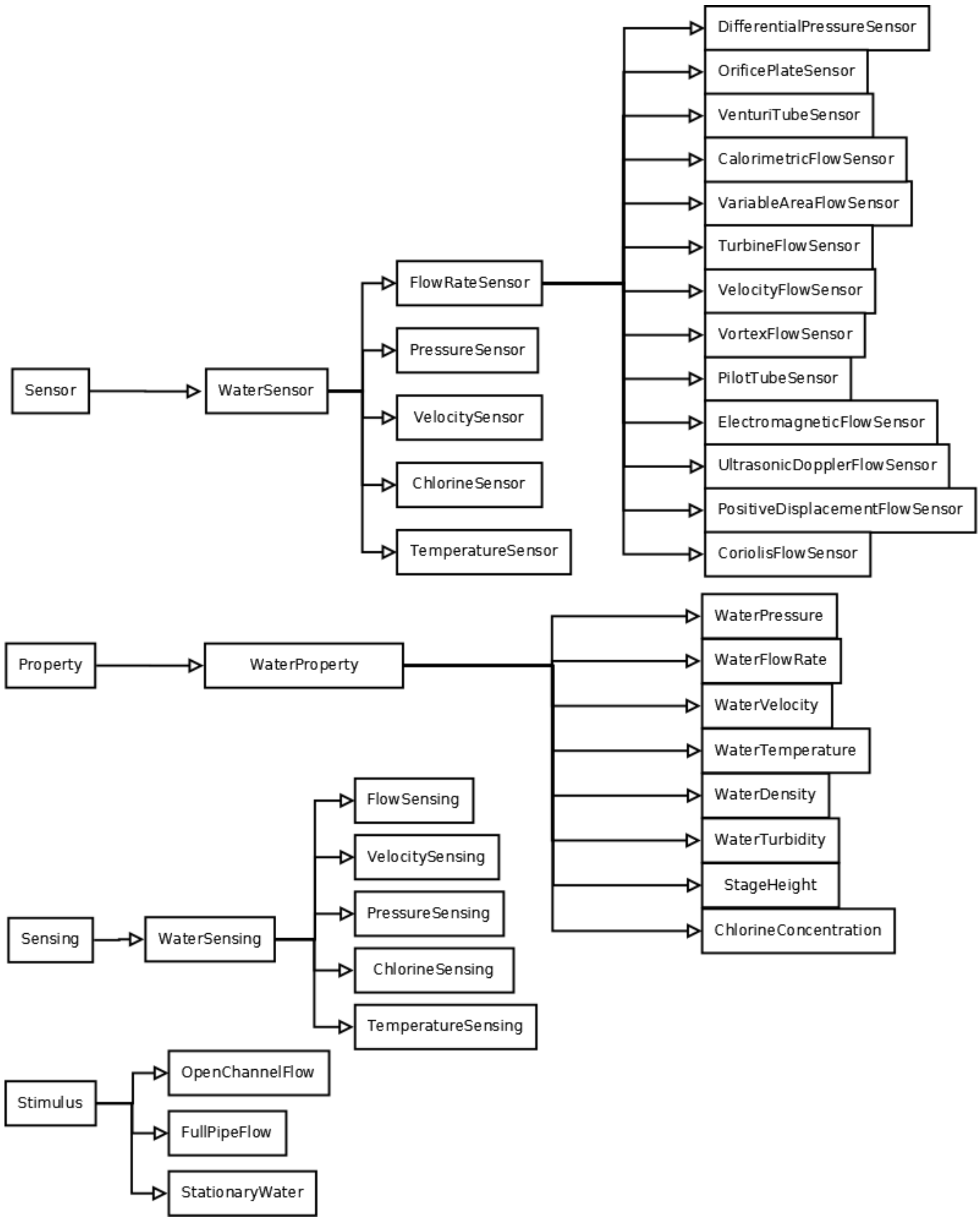


**Figure 51: The main class hierarchy extensions to the SSN ontology for WISDOM**

## 4.3.4.3.1 DATA ENRICHMENT & WCIM LINKS

Relationships between the cyber and physical concepts in the domain were modelled. The key relevant relationships are shown in Figure 52 (inverse relationships not shown for simplicity).

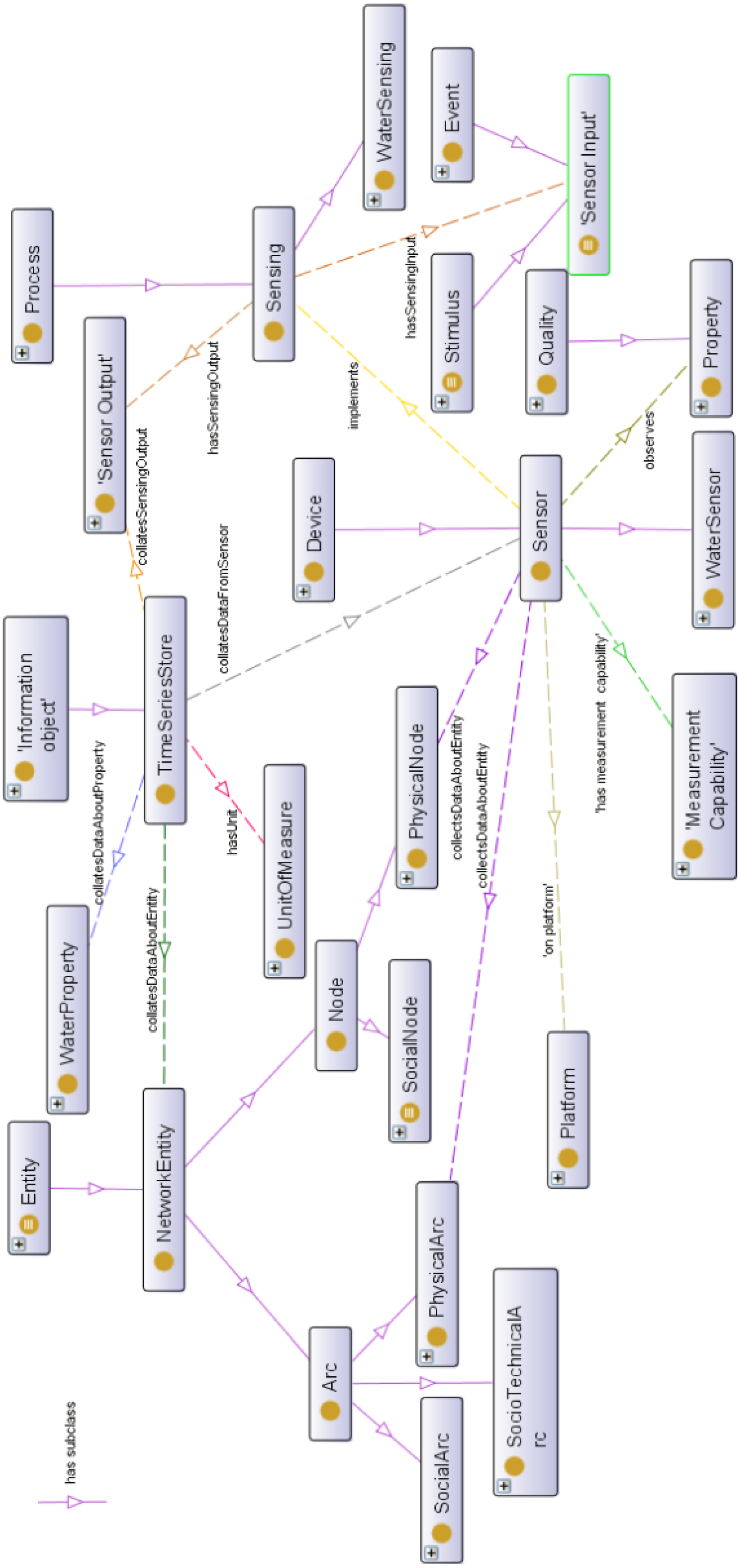


Figure 52: Key relationships between WCIM & Sensor Ontology concepts

Through the relationships shown in Figure 52, data have been contextualised in terms of the component they relate to, their unit, the sensor they originated from and the water property they relate to. This was aligned with the semantics of the KairosDB module, as shown in Appendix A.

#### 4.3.4.3.2 PROBLEMS AND ALERTS

Based on industry feedback, alarms, problems, and alerts, were also modelled, as shown in Figure 53. This allowed the developed rule engine to provide more powerful inference. Entity-problem relationships are inferred at runtime based on sensor data. Further, WaterAlert individuals have an '*isActive'* Boolean property. As description logic is monotonic, semantic reasoning cannot infer the existence of new individuals. This means that the alerts and problems must either all be defined in the ontology prior to deployment, or external software must create new named individuals when a problem arises. A simple 'acceptable range' condition has been modelled, as this serves as proof of concept.

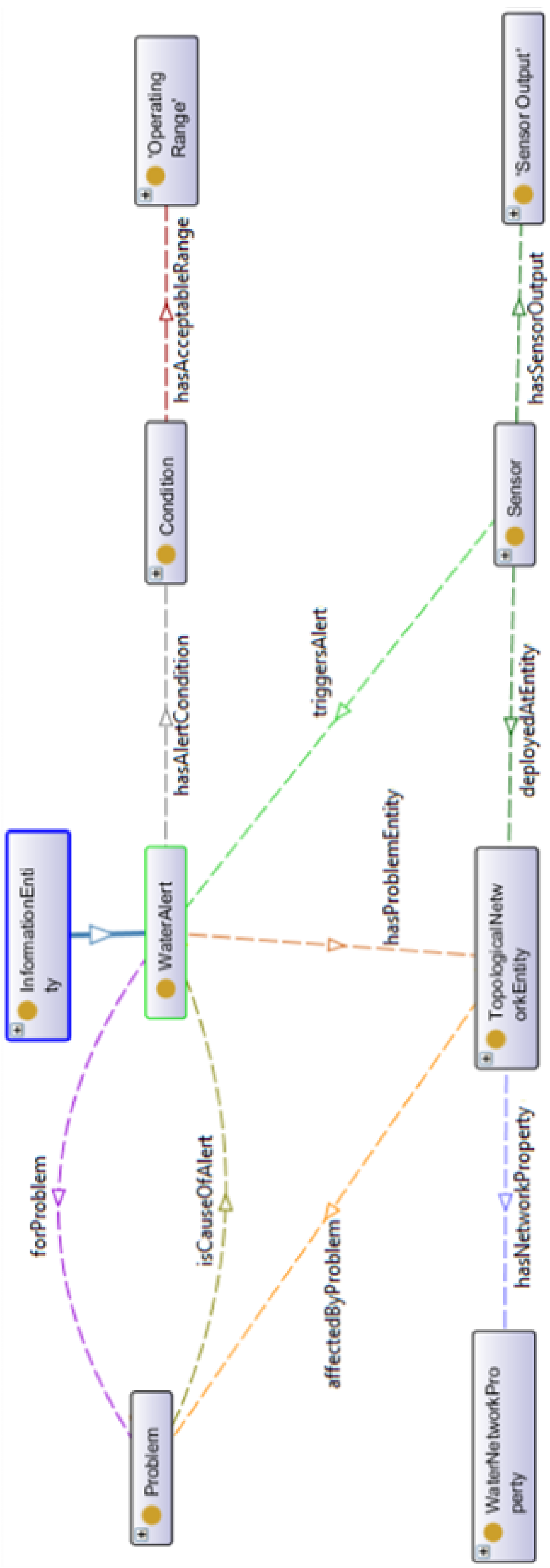


**Figure 53: Key concepts and relationships regarding network problems and alerts**

#### 4.3.4.4 WATER VALUE CHAIN SOCIAL MODEL

The Water Value Chain Social Model (WVCSM) formalizes the domain vocabulary for the social nodes and arcs of a water network. In contrast to the physical model, social arcs are modelled as classes themselves as relationships between social entities are more often referred to than between physical entities, such as describing the details of a contract.

Social networks only include nodes connected by arcs. Social nodes are referred to as ‘agents’, which are typically people or organisations. The main class hierarchy of the social model is shown in Figure 94. Again, data properties describe the organisations and people themselves as appropriate.

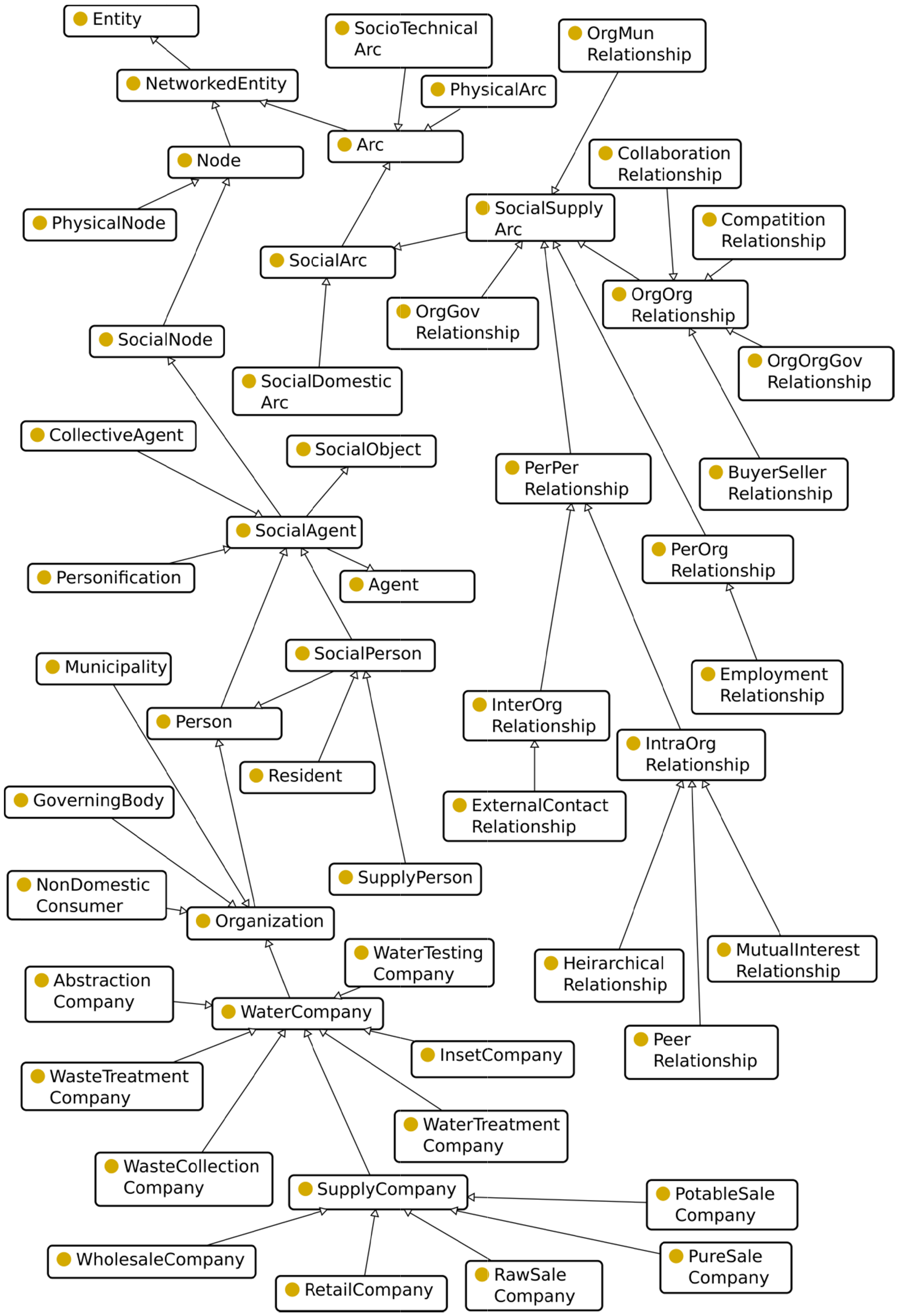


**Figure 54: WVCSM class hierarchy of main social network entities**

A social 'node' is connected to any number of arcs, and a social arc connects 2 nodes. Supply side social arcs include employment contracts, regulations and management structure relationships. An example social graph is shown in Figure 55. Each of the node and arc classes shown in Figure 55 has a taxonomy of sub-classes.

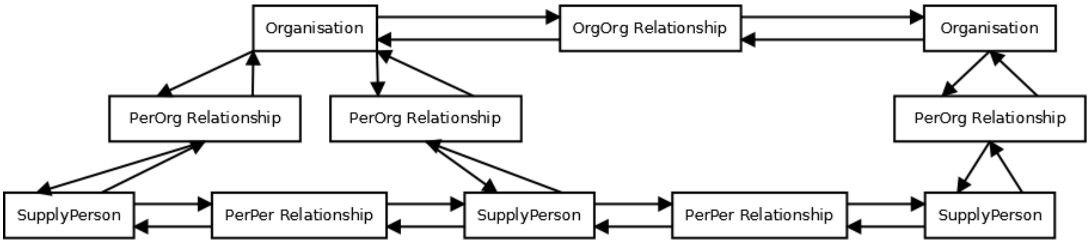


**Figure 55: Example supply-side social network**

This model uses a system-of-systems approach, where knowledge can be captured both at the interpersonal and/or inter-organisational level. This represents a multiple level-of-detail approach.

The supply-side (clean & waste services) social model and the domestic social model serve significantly different purposes. The supply-side model enriches asset descriptions and sensory data and facilitates knowledge management, whereas the domestic model categorises a 'domicile' based on the social system which occupies it and its water use. The domestic social model is further enriched with demographic and water data and object properties, more so than other areas of the ontology.

#### 4.3.4.4.1 WATER MANAGEMENT STAKEHOLDERS

The main stakeholder organisations within a water value chain have been categorised as shown in Figure 56. Individual organisations could be an instance of multiple subcategories, such as a supply company being both a '*RetailCompany'* and a '*PotableSaleCompany'*.

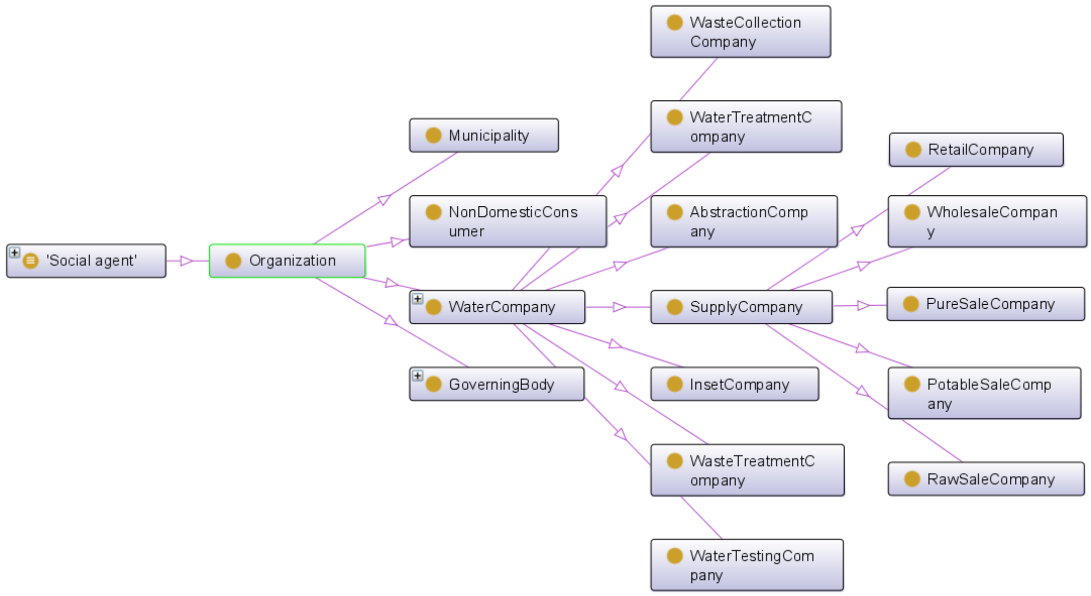


**Figure 56: WVCSM water management organization class hierarchy**

## 4.3.4.4.2 SUPPLY-SIDE RELATIONSHIPS

Supply-side relationships are modelled as entities in themselves so that they can be described in detail. These relationships are categorised as either person-person, person-organisation, organisation-organisation, organization-governing body or organization-municipality; where 'organisation' refers to a water company. The main class hierarchy is shown in Figure 57.

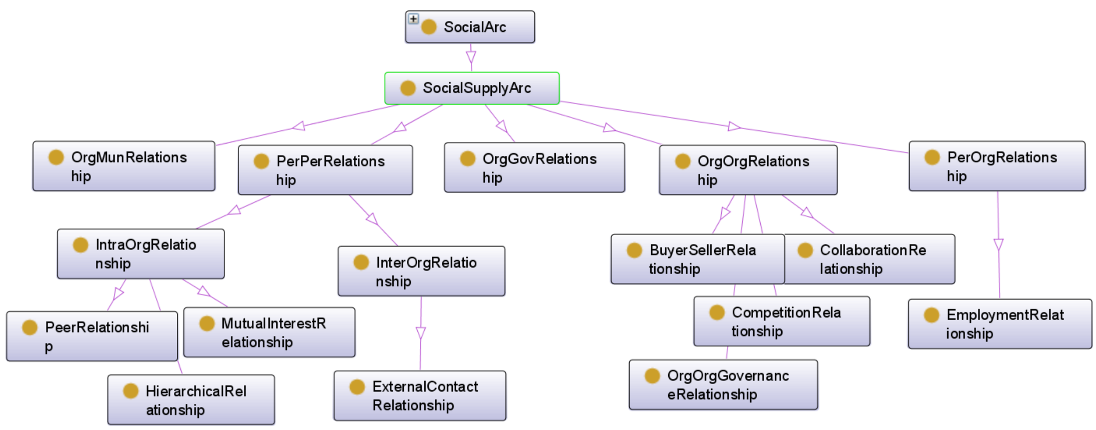


**Figure 57: WVCSM class hierarchy of supply side relationships**

### 4.3.4.4.3 DOMESTIC SOCIAL RELATIONSHIPS

To classify domestic consumption profiles and formalise domestic level water use behaviour, domestic social networks were modelled. Relationships and people can be described through data properties and supplementary object properties if necessary for target applications. The main relationships are family and neighbourly relationships, which can broadly categorise domiciles into usage profiles, as shown in Appendix A.

### 4.3.4.4.4 SUPPLY SIDE SOCIO-TECHNICAL RELATIONSHIPS

Relationships between social, physical and virtual entities of the water value chain were modelled, as shown in Figure 59. Again, these are categorised as existing either at the supply or domestic level. Whereas physical relationships were modelled as object properties and social relationship types were modelled as classes, socio-technical relationships were modelled as both classes and object properties. It is variable whether the relationship itself will require much description, and directly relating instances is typical practice. The nature of these socio-technical classes is shown in Figure 58 below with an example on both the supply and domestic side.

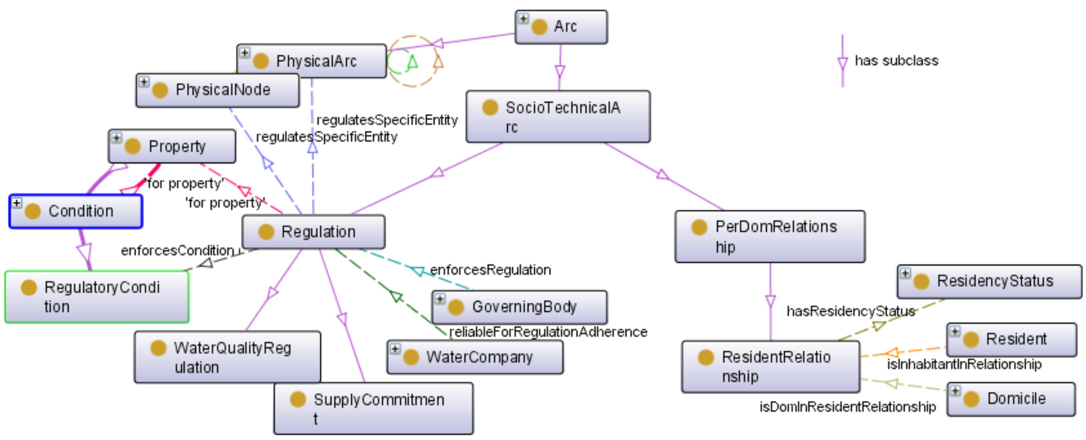


**Figure 58: WVCSM socio-technical arc classes on both supply and demand side**

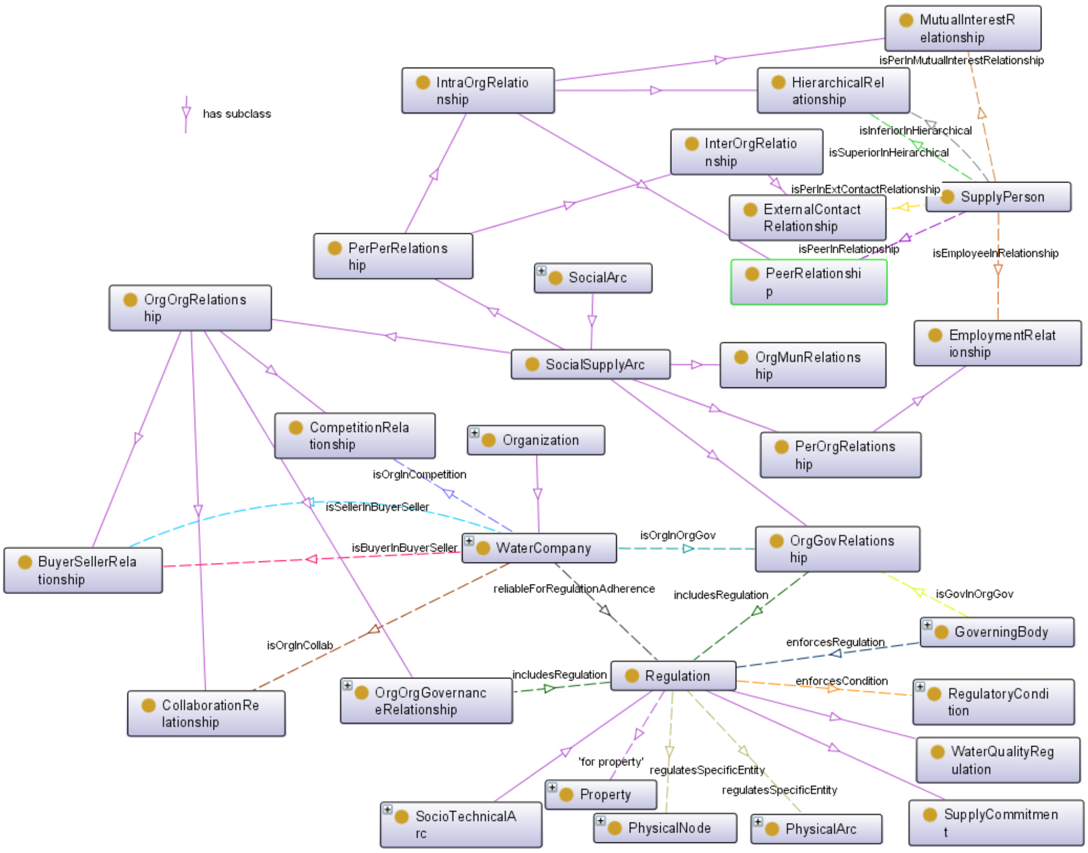


**Figure 59: WVCSM main supply side social and socio-technical classes and relationships**

### 4.3.4.4.5 DOMESTIC SOCIO-TECHNICAL RELATIONSHIPS

Domestic socio-technical water systems were simplified to only model consumption and conservation behaviours, ignoring domestic plumbing and flow control. An '*ApplianceConsumptionPattern'* is introduced, as shown in Figure 60. The notion of inhabiting a domicile has also been modelled.

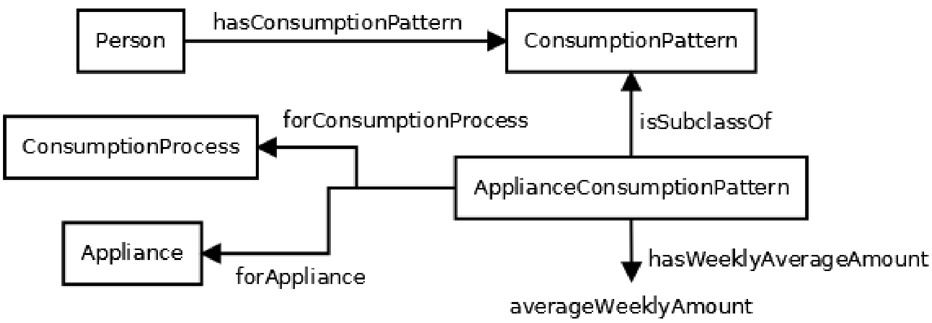


**Figure 60: WVCSM Consumption pattern concept and relevant object properties**

#### 4.3.4.4.6 ECONOMIC CONCEPTS & RELATIONSHIPS

Some economic concepts have been included, such as the domestic water bill; this is related to a person through the '*main bill payer*' relationship and to a tariff through the '*charged according to*' relationship. The tariff could be a static rate or could be a dynamic pricing scheme. Economic concepts at the supply level were out of scope.

### 4.3.5 ONTOLOGY VALIDATION

The initial, automated check of the ontology's consistency through the built in Protégé reasoner has consistently passed many times throughout the ontology's development; such that the ontology does not contain contradictory statements. The competency questions have been utilized throughout the semantic activities, and they can be answered by the ontology. This has been checked thoroughly by 'asking' the questions as SPARQL queries once the pilot site knowledge bases were finalized. The domain expert validation was conducted separately with the Italian and the Welsh domain expert partners through one day workshops, and in both cases the ontology's modelling choices were broadly validated, the majority of the detailed modelling choices were validated and corroborated between workshops, and some revisions and extensions were suggested. These modifications have been made and are reflected in the version of the model presented in this thesis. An additional workshop with the WISDOM partners and special interest group experts was then conducted, which served to validate that the changes made were sufficient and hence that the ontology is now sufficient. This workshop was held with parties from all WISDOM pilot sites present, so as to produce discussion and a consensus of either validity, or of the necessary revisions.

The domain ontology was tested for validity at a project-facilitated meeting of industrial experts, where it was considered by a wider range of stakeholders in the water value chain, most of which had little bias towards the WISDOM project. This offered a broad view on the ontology and hence tested its extent, as well as its detail in areas of the water value chain which the WISDOM partners are not experts in. That consensus was reached that the WISDOM ontology represents a shared and sufficient conceptualization of the domain by this group, represents a significant milestone in its validation. Finally, the domain ontology was tested in an integrated manner with the other semantic components through the web service testing, which determined that the ontology successfully and sufficiently contributes to an 'ICT for water' web platform. This integration with the other WISDOM components has been conducted to a preliminary stage, and represents ongoing work within other project tasks.

Some of the comments from the SIG expert validation session were:

1. The ontology addresses the problem of interacting between tools (GIS, SAP, customer data)
2. Include alarms as well as sensors
3. 'Governing body' is also called 'regulator'
4. Include 'water testing company'

The 2$^{nd}$ comment is addressed in Section 4.3.4.3.2, the 3$^{rd}$ comment has been addressed by adding a comment to the class, and the 4$^{th}$ comment was addressed by including a 'waterTestingCompany' class. The majority of comments were advisory or generic, such as regarding possible future work, rather than required changes in the scope currently addressed. Examples of these comments were:

- The work could be considered as a type of enterprise service bus
- An ontology is also called a taxonomy
- Sensors could also be 'social sensors', which report numbers of tweets etc.
- Collaboration relationships exist between utilities which share a water resource

Table 29 below presents an example outcome of the competency question testing, showing how the deployment sufficiently answers the questions when formalized as SPARQL queries, where the queries were answered in circa 15ms. Overall, the ontology has been validated by 25 organisations, with a range of expertise, as illustrated in Figure 61.

**Table 30: Example competency question testing evidence**

| |
|---|
| Natural language question: |
| What is sensor E2000’s current reading? |
| SPARQL query |
| PREFIX wis:<http://www.WISDOM.org/WISDOMontology#> |
| PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> |
| PREFIX dul: <http://www.loa-cnr.it/ontologies/DUL.owl#> |
| |
| SELECT ?reading |
| WHERE { |
| wis:E2000 rdf:type wis:LevelSensor. |
| wis:E2000 wis:hasLatestOutput ?output. |
| ?output dul:hasDataValue ?reading } |
| Output (csv format) |
| reading |
| 2.0 |
| Natural language question: |
| What is Pipe X’s material? |

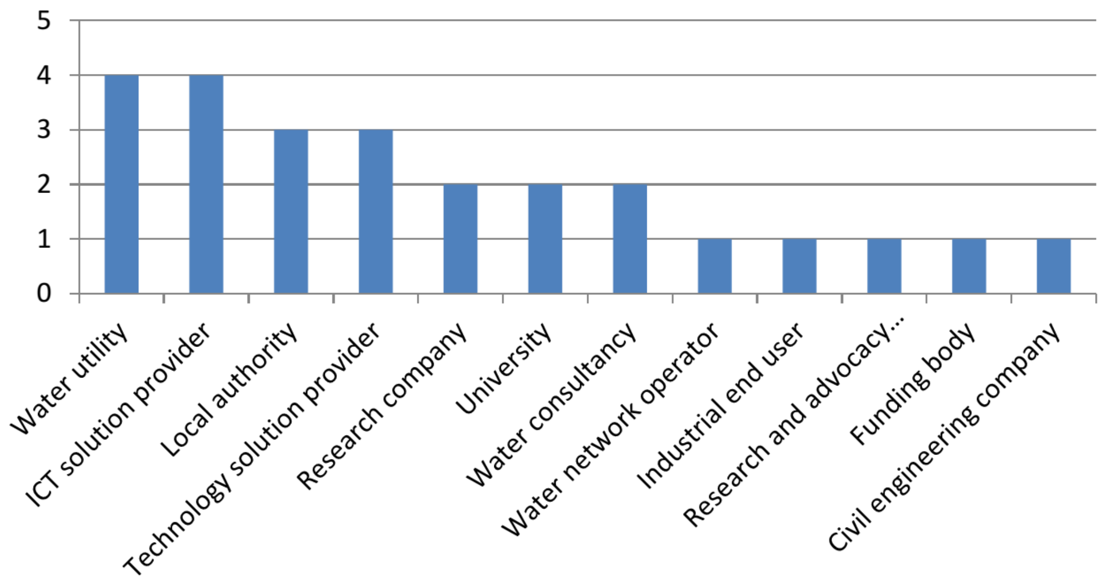


**Figure 61: Breakdown of ontology validators by organisation type**

## 4.3.6 ALIGNMENT WITH EXISTING STANDARDS AND MODELS

### 4.3.6.1 ALIGNMENT WITH WATERP ONTOLOGY

The WISDOM ontology was aligned to the WatERP ontology, and hence with the recommendations of WaterML2. This was stated through OWL annotations for relevant classes and properties.

The majority of modelling patterns adopted by the two ontologies were homogeneous from their joint grounding in the SSN ontology. Further matching was achieved through manual alignment, which involved revising the WISDOM ontology by adding concepts and axioms and revising existing ones to produce a compatible domain perspective which still met the requirements. The matched terms are stated in Table 49 in Appendix A.

A system modelled through the WatERP ontology could be extended with the WISDOM ontology's detailed physical network concepts without having to redo or change the existing knowledge base. Some of the WatERP modelling patterns and concepts could not be aligned with the WISDOM ontology, as shown in Table 50, although they are still complementary. A WatERP pipe could be described using WISDOM concepts, such as in the triple WatERP:Pipe_001 RDF:type WIS:TrunkMain. This would then allow a reasoner to infer extra knowledge about the pipe based on the WISDOM Tbox, and would allow ontology-driven applications to leverage both domain perspectives on the available Abox.

The key areas of the WatERP ontology which are outside of WISDOM's scope are FinanceManagementFlow, Instruments, and AssessmentIndicator. The main aspects of the WISDOM ontology which are out of WatERP's scope are WasteNetwork concepts, ConsumptionProcess subclasses, DesignedArtefact subclasses, DesignedArtefact properties, Sociotechnical relationships, Sensor subclasses, MeasurementCapability, Sensing, and subclasses and object properties which model additional depth to the concepts modelled by WatERP.

### 4.3.6.2 ALIGNMENT WITH THE INDUSTRY FOUNDATION CLASSES

Broad alignment with the non-ontological resource of the relevant parts of the Industry Foundation Classes (IFC) was achieved, as shown in Table 51 in Appendix

A, although the IFC schema assumes that all individuals are within the context of a building, or at least an 'architecture, engineering and construction' scope, so further testing and refinement is needed for this alignment.

#### 4.3.6.3 ALIGNMENT WITH INSPIRE

The INSPIRE directive [512] aims to facilitate geospatial data exchange within Europe, so as to foster international coordination in situations such as river contamination (where rivers cross country borders), and so has produced a set of data specifications, as UML class diagrams. These artefacts (from their application theme of utility networks and government services) include a data specification for generic networks and utility networks, and slightly more specific models for sewer networks and water networks. These model high level topological relationships and entities such as nodes and arcs, as well as some enumerations such as water types, node types, and warning types. Whilst full alignment with the code lists which INSPIRE propose is out of scope, this could be achieved given the alignment of critical modelling decisions. This alignment is shown in Table 52 in Appendix A as a set of aligned terms, which represent classes and object properties.

#### 4.3.6.4 PLACEMENT WITHIN EXISTING STANDARDS

Based on the reuse and alignments of the WISDOM ontology with other knowledge modelling artefacts and standards, it is well situated to serve a role in the standardisation landscape, following some development of its maturity through industrial exposure and accepted standardisation processes. The ontology builds on existing work and interoperates with relevant standards in the water, IoT, and knowledge management fields, as illustrated in Figure 62.

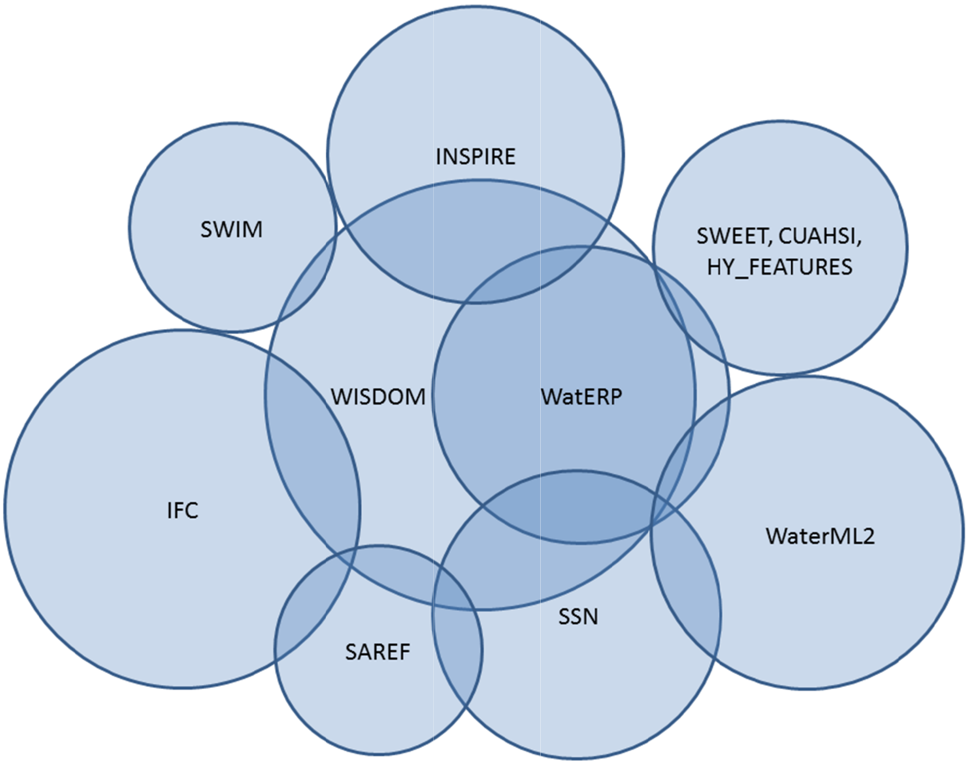


**Figure 62: Alignments between WISDOM ontology and other models and standards**

### 4.3.7 INSTANTIATION OF KNOWLEDGE BASES

After the domain ontology (Tbox) development, knowledge bases (Tbox + Abox) were instantiated at each of the 4 pilot site. This work produced a novel Python script for converting GIS data to RDF data. The section briefly discusses the Tbox – Abox approach further, before describing the scripts and knowledge bases produced.

#### 4.3.7.1 KNOWLEDGE BASE INTRODUCTION

The semantic knowledge bases produced here continue the use of the W3C semantic web framework adopted in formalising the ontology as an OWL ontology. The domain ontology presented represents a language which can be used to describe a water value chain, and each knowledge base uses this language to describe a pilot site. In this manner each pilot site knowledge base uses the same language, but is itself entirely separate to the other pilot site knowledge bases. That is not to say that knowledge cannot be shared between pilot sites, just that they have been separated for manageability, and because little direct utilisation of each other's data was undertaken.

Each knowledge base was produced by reusing existing data from utility company project partners where possible, and completing the remaining instantiation manually, to the completeness required in order to satisfy the requirement specification. This process is now described, before each of the knowledge bases is presented in turn.

#### 4.3.7.2 KNOWLEDGE BASE RESULTS

This section presents the pilot site models developed, following the methodology described in the previous subsection. The input and output data for each model is summarised, before the main modelling patterns used and objects modelled in the knowledge bases are illustrated.

7 CSV files were used as a basis for extracting information regarding the Cardiff pilot site; these described the system valves, meters, mains, control valves, boundary valves, hydrants, and asset sensors. In total, these represented 352 KB of data. The number of entities and properties in each of these sheets is summarised in Table 53 in Appendix B.

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 2183 KB in size, representing 17355 triples. This included the 1966 named entities, and the properties shown in Table 54 in Appendix B.

6 CSV files were used as a basis for extracting information regarding the Tywyn and Aberdovey pilot site; these described the system valves, meters, mains, control valves, hydrants, and asset sensors. In total, these represented 572 KB of data. The number of entities and properties in each of these sheets is summarised in Table 55 in Appendix B.

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 2142 KB in size, representing 21122 triples. This included the 2363 named entities, and the properties shown below in Table 56 in Appendix B.

As the Gowerton pilot site only includes waste water assets, its knowledge base is significantly different to the Cardiff and Tywyn & Aberdovy sites. 5 CSV files were used as a basis for extracting information regarding the Gowerton pilot site; these described the conduits, nodes, pumps, sensors and subcatchments. In total, these represented 5.28 MB of data. The number of entities and properties in each of these sheets is summarised in Table 57 in Appendix B.

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 6582 KB in size, representing 60081 triples. This included the 6674 named entities, and the properties shown below in Table 58 in Appendix B.

The Italian Pilot site was instantiated using a Python script to extract knowledge from the hydraulic model developed of the Italian pilot site, using the RDFlib Python library previously described and the EPANETTOOLS Python library [571]. This meant that instead of the Welsh pilot method of exporting CSV files, then parsing the CSV files into Python objects, the EPANET model could be parsed directly into Python objects, and then iterated over to add statements to an RDFlib graph and the WISDOM namespace, before being serialized into an RDF file. The Italian pilot input data was split across several sections of an EPANET input file, which described the 426 entities within 42 KB of data, as detailed in Table 59 in Appendix B.

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 222 KB in size, representing 1942 triples. This included the 426 named entities, and the properties shown below in Table 60 in Appendix B.

### 4.3.8 SEMANTIC WEB SERVICE TESTING

#### 4.3.8.1 SOFTWARE PERFORMANCE

Following the validation of the domain ontology, this was instantiated for a real Welsh pilot site by using survey data from residents as well reusing GIS data and data from sensor, social and asset databases, as well as heuristic knowledge, operating manuals and product specification sheets. This pilot site knowledge base was then deployed in the cloud based system described previously; with live data updating the instantiation every 15 minutes. Testing was conducted as to the performance of the ontology service within the cloud platform for both retrieval and updating of data, through the RESTful GET and PUT methods. These utilized the SPARQL SELECT and UPDATE functions respectively.

The service was deployed on a personal laptop (i5-3317U CPU@1.7GHz, 8GB memory, Windows 7 64-bit) so as to test the service's performance, rather than including latency by testing the service in a cloud environment. The semantic model tested was an instantiation of the water value chain and domestic model, consisting of 1722 named individuals. 11 identical GET requests were issued to the service to retrieve the current sensor reading at an arbitrary sensor in the network, and this test was repeated 5 times, with the service restarted between each test to reset any caching which had occurred. A similar testing protocol was conducted for PUT requests to update the sensor reading, and more realistic testing was conducted by varying the GET request issued, varying the PUT request issued, and finally alternating between GET and PUT requests. The results of the GET request testing are shown in Figure 63 below, which clearly shows caching, and that the typical response time which could be expected would be circa 550ms. The PUT testing indicated a very similar trend, but with approximately an additional 100ms response time across the requests. Changing the request between subsequent requests didn't result in any significant difference in the response time to these results.

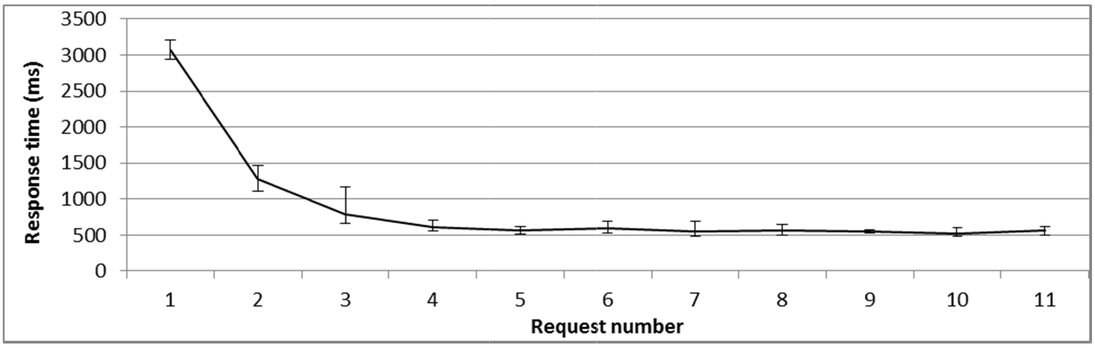


**Figure 63: Average response time of the ontology web service across several SELECT queries**

The ontology service consumed circa 113MB of memory on start-up, and following caching, peaked at circa 800MB after 20 requests. Following start-up, a request consumed on average 81% of the available processing power, but after 5 requests this reduced and stabilized at circa 11%.

#### 4.3.8.2 SMART HOME APP SCHEMA CONVERSION

The benefit of using the semantic web approach to promote interoperability across software with different domain perspectives was tested by performing a schema conversion from a knowledge base of devices instantiated within the WISDOM ontology into a set of SAREF individuals. This RDF data could then be used within an UPDATE SPARQL query to add the individuals to a SAREF knowledge base. A similar approach could be used in the more likely case of converting to an application specific ontology which is also mapped to the SAREF ontology, by loading both ontologies and the SAREF ontology into memory. This could also convert object and data properties between knowledge bases if appropriate mappings were formalized. The conversion was conducted through a simple SPARQL CONSTRUCT query. Excerpts of the source data, SPARQL query and output data are shown in Table 30 below, in turtle. Careful federation of the shared objects would be required to manage access rights and update priority, for example whether the application using the target knowledge base could update properties regarding individuals at the source knowledge base URI. The implication of this is that software developers could utilize data from across these domains far more easily, more powerfully, and with more confidence that the data was being correctly understood and contextualized.

**Table 31: Excerpts of the knowledge base conversion process, with prefix definitions omitted**

| Source data |
|---|
| wisdom:washingMachine rdf:type wisdom:ElectricAppliance, owl:NamedIndividual<br><br>wisdom:meter_01 rdf:type wisdom:DomesticWaterMeter, owl:NamedIndividual<br><br>wisdom:meter_02 rdf:type wisdom:DomesticWaterMeter, owl:NamedIndividual |
| SPARQL query |
| CONSTRUCT { ?individual rdf:type owl:NamedIndividual, ?TargetName} |

| |
|---|
| WHERE{ ?individual rdf:type owl:NamedIndividual.<br>?individual rdf:type ?SourceName.<br>?SourceName wisdom:alignedWithSaref ?TargetName<br>} |
| Output data |
| wisdom:washingMachine rdf:type saref:Device, owl:NamedIndividual<br>wisdom:meter_01 rdf:type Saref:Meter, owl:NamedIndividual<br>wisdom:meter_02 rdf:typeSaref:Meter, owl:NamedIndividual |

### 4.3.9 ADVANCED DECISION SUPPORT THROUGH INFERENCE AND SEMANTIC RULES

Following positive feedback from practitioners, a knowledge-based system was developed by coupling the triple store with an SWRL rules and an inference engine, and producing a graphical interface. This section introduces the target use case and software architecture, the semantic inference rules produced, and finally the graphical interface.

#### 4.3.9.1 OVERVIEW AND USE CASE DESCRIPTION

A new use case was produced specifically to highlight the value of semantic inference. The proposed use case assumed a fault had occurred within the water value chain, such as a pipe blockage. The aim of the target system was to assist the decision maker in responding to this issue. A core software requirement was that the inference engine should have the ability to detect problems in the network, and then determine the network entities affected by the problem. This would provide a decision support function, for example helping to identify customers affected by a network blockage and proactively engaging with them.

Identification of faults was achieved through rule-based detection. The concept of an alert was modelled in the domain ontology, as well as the problem which caused the alert, and object properties then connected these to the physical network entities, their topological representation, and the related sensors and properties. Each alert had an associated '*alert condition'*, which could be a complex fault detection algorithm, but a simple '*acceptable range'* was used for proof of concept. This range had an upper and lower bound, which an SWRL rule was able to evaluate against the latest observation from the appropriate sensor. Firstly, the rule engine determined the affected network entities, and its severity and detection time. By updating the knowledge base with this information, it was then exposed to applications.

The work used KairosDB as a timeseries database, and the Pellet reasoner for simpler rules and OWL-based inference, with a separate Drools engine managing more complex rules. The Pellet reasoner was chosen over native Jena reasoners to achieve the maximum reasoning capabilities from native OWL axioms. The Drools engine was found to offer better performance and reliability during testing. At each timestep, the Drools engine re-evaluates the triple store against the rules, and updates the triple store accordingly.

#### 4.3.9.2 INFERENCE ENGINE RULES

The required inference capability for the target use case is illustrated in Figure 64, where the solid arrows indicate explicit knowledge, and dashed arrows represent inferred knowledge.

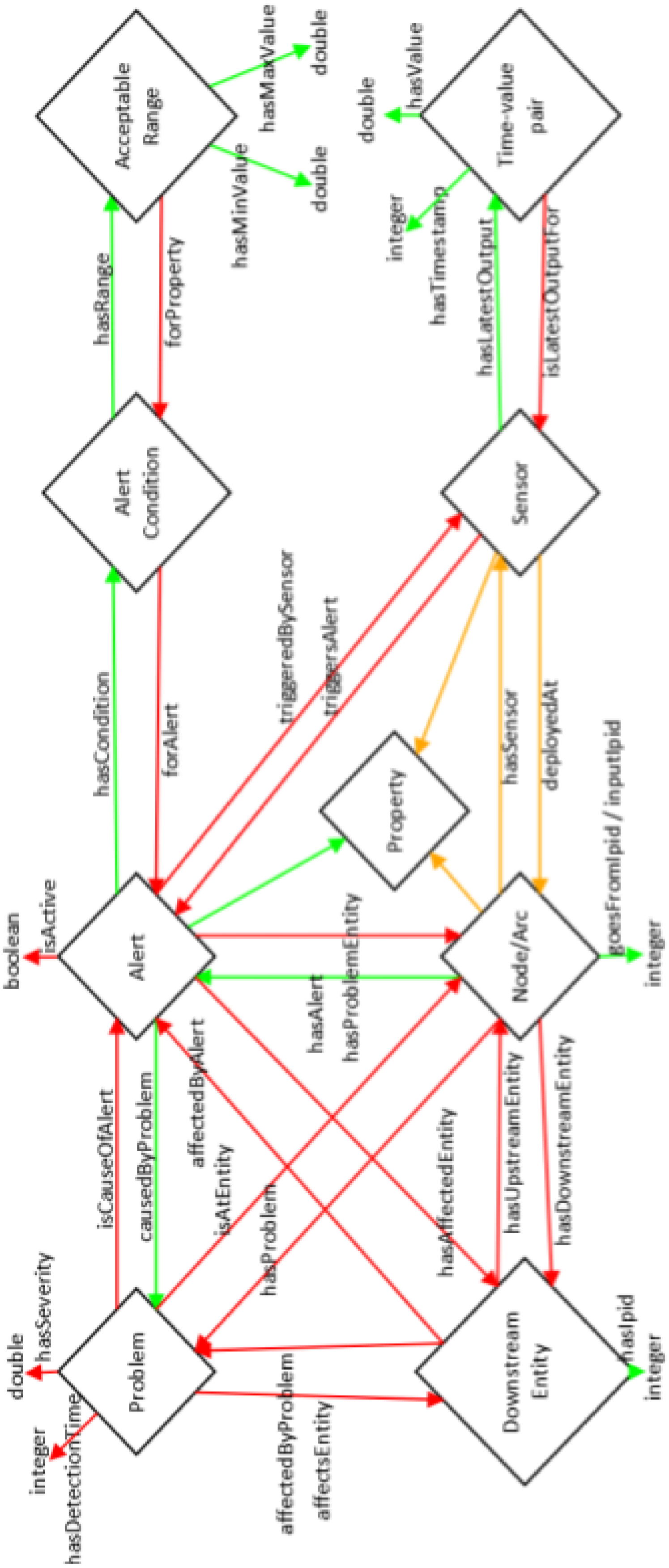


**Figure 64: Problem detection and alert propagation**

A number of SWRL rules were written to achieve the desired inference, and were utilized by the inference engine. Example rules are now described and presented in SWRL syntax, and the full set of rules can be found in Appendix C.

**Inferred property: deployedAt**

As sensors are not explicitly described in terms of the node which they are deployed at, this is fundamental knowledge which must be inferred in order to contextualise the capability of the deployed sensors.

```
SENSOR(?S) ^ ATASSET(?S, ?A) ^ TOPOLOGICALNETWORKENTITY(?E) ^ ATASSET(?E, ?A) -> DEPLOYEDATENTITY(?S, ?E)
```

**Inferred property: isActive**

If an alert has an acceptable range, and it is triggered by a sensor, and the sensor's latest reading falls outside of the acceptable range, the alarm is triggered. This can occur if the reading is greater than the allowable maximum, or smaller than the allowable minimum, otherwise the alarm is not active.

```
WATERALERT(?A) ^ HASALERTCONDITION(?A, ?AC1) ^ HASACCEPTABLERANGE(?AC1, ?AR) ^ HASMAXVALUE(?AR, ?XMAX) ^ SENSOR(?S) ^ TRIGGERSALERT(?S, ?A) ^ HASLATESTOUTPUT(?S, ?TVP) ^ HASVALUE(?TVP, ?X) ^ SWRLB:GREATERTHAN(?X, ?XMAX) -> ISACTIVE(?A, TRUE)
```

```
WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:LESSTHAN(?X,?XMIN)-> ISACTIVE(?A,TRUE)
```

```
WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:LESSTHANOREQUAL(?X,?XMAX)^SWRLB:GREATERTHANOREQUAL(?X,?XMIN)-> ISACTIVE(?A,FALSE)
```

**Inferred property: hasDownstreamEntity**

In order to generalise pipes, pumps, and reservoirs etc. to determine what is upstream or downstream of an entity, it is useful to use the IPID values held in the legacy GIS database to infer knowledge about flow chronology through the entities. This allows later inference of whether an entity is affected by any given problem, and greatly simplifies those rules.

If an entity goes from an entity with IPID of i, and another entity has an IPID of i, then the latter is downstream of the former, and vice versa.

GOESFROMIPID(?P, ?I) ^ HASIPID(?U, ?I) -> HASUPSTREAMENTITY(?P, ?U) ^ HASDOWNSTREAMENTITY(?U, ?P)

GOESTOIPID(?P,?I)^HASIPID(?D, ?I) -> HASUPSTREAMENTITY(?D, ?P) ^ HASDOWNSTREAMENTITY(?P, ?D)

**Inferred property: hasDetectionTime**

Given that the knowledge base will be iteratively updated as new sensor readings are received, and alerts may not be observed immediately, it would be beneficial to inform decision makers exactly when a problem was first observed. This is achieved by noting the time at which the sensors latest reading is outside the acceptable range, but the sensor's previous reading was inside the acceptable range.

If an alert has an acceptable range, and is triggered by a sensor, and the sensor's latest reading is outside that range, but its previous reading was inside the range, then the detection time of the problem is the latest reading's timestamp. This can occur when the reading is above the maximum range, or below the minimum range.

PROBLEM(?P)^ISCAUSEOFALERT(?P,?A)^WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:GREATERTHAN(?X,?XMAX)^HASTIMESTAMP(?TVP,?TIME)^HASPREVIOUSOUTPUT(?S,?TVPPREV)^DIFFERENTFROM(?TVP,?TVPPREV)^HASVALUE(?TVPPREV,?XPREV)^SWRLB:LESSTHANOREQUAL(?XPREV,?XMAX)^SWRLB:GREATERTHANOREQUAL(?XPREV,?XMIN)->HASDETECTIONTIME(?P,?TIME)

PROBLEM(?P) ^ ISCAUSEOFALERT(?P, ?A) ^ WATERALERT(?A) ^ HASALERTCONDITION(?A, ?AC1) ^ HASACCEPTABLERANGE(?AC1, ?AR) ^ HASMINVALUE(?AR, ?XMIN) ^ HASMAXVALUE(?AR, ?XMAX) ^ SENSOR(?S) ^ TRIGGERSALERT(?S, ?A) ^ HASLATESTOUTPUT(?S, ?TVP) ^ HASVALUE(?TVP, ?X) ^ SWRLB:LESSTHAN(?X, ?XMIN) ^ HASTIMESTAMP(?TVP, ?TIME) ^ HASPREVIOUSOUTPUT(?S, ?TVPPREV) ^ DIFFERENTFROM(?TVP, ?TVPPREV) ^ HASVALUE(?TVPPREV, ?XPREV) ^ SWRLB:LESSTHANOREQUAL(?XPREV, ?XMAX) ^ SWRLB:GREATERTHANOREQUAL(?XPREV, ?XMIN) -> HASDETECTIONTIME(?P, ?TIME)

### 4.3.9.3 INFERENCE USE CASE TESTING

The rules were tested individually during development for efficacy and to stimulate false positives, which the rule set consistently passed. However, use case based testing of the rules in unison is more rigorous. This section therefore presents the use case based testing of the rules. The rules were tested on a Samsung 900X laptop, with an Intel i5 1.7GHz processor and 8GB of memory, on 64-bit Windows 7, and the rules were tested in the Protégé software. The rules were tested on an instance of the domain ontology, such that the entire domain ontology was reasoned over, as well as the individuals specifically relevant to the use case. The input and output graph sizes were determined through the RDFlib Python library by command line.

An instance of the use case was defined whereby a reservoir node is connected to a level sensor, and has a tree of downstream nodes and arcs. It was considered that the sensor's latest reading indicated that the reservoir's water level was too low, as illustrated in Figure 65, which also shows the named individuals for the alert, and acceptable range and latest output.

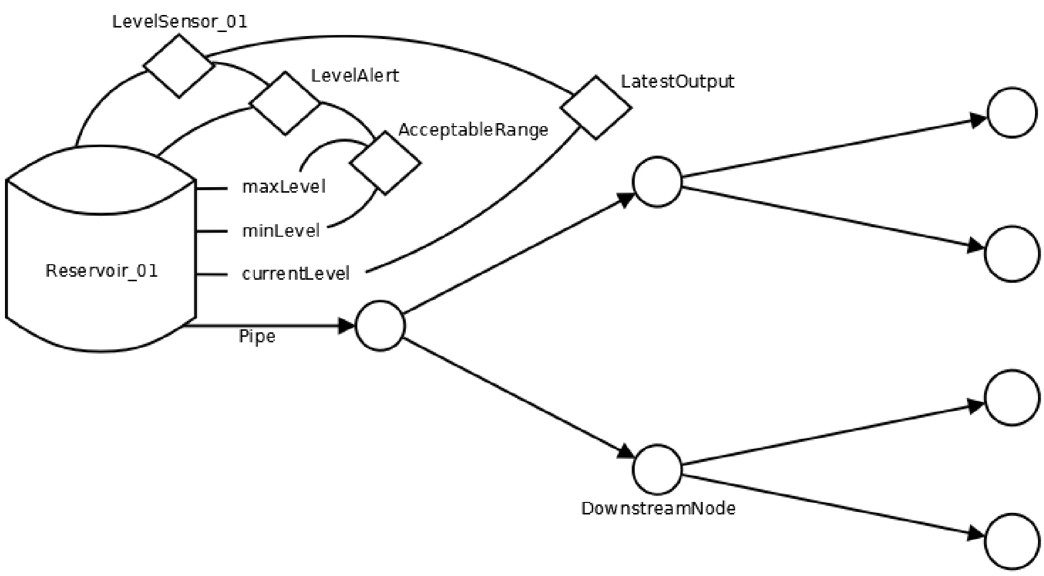


**Figure 65: Problem and alert inference test case illustration**

Following the application of the inference engine on the test knowledge base of 56 triples, the Abox contained 972 triples, meaning that 916 triples had been inferred, of 7185 total inferred axioms. This inference occurred in 1427 ms on the first instance (without caching), which reduced to circa 450 ms after caching. The desired knowledge primarily centres on the problem and downstream entity named individuals, so Figure 66 displays the Abox knowledge at these entities following the inference. This shows that the node is linked to its upstream and downstream entities, is ‘affected by’ the active alert, and is ‘affected by’ the low level problem. Figure 66 also shows that the problem individual is linked to all of the downstream nodes, and its severity and timestamp have been inferred.

Figure 67 highlights some of the key inferred knowledge. Specifically, this shows that knowledge about the reservoir problem can be inferred, and this can be linked directly to downstream entities.

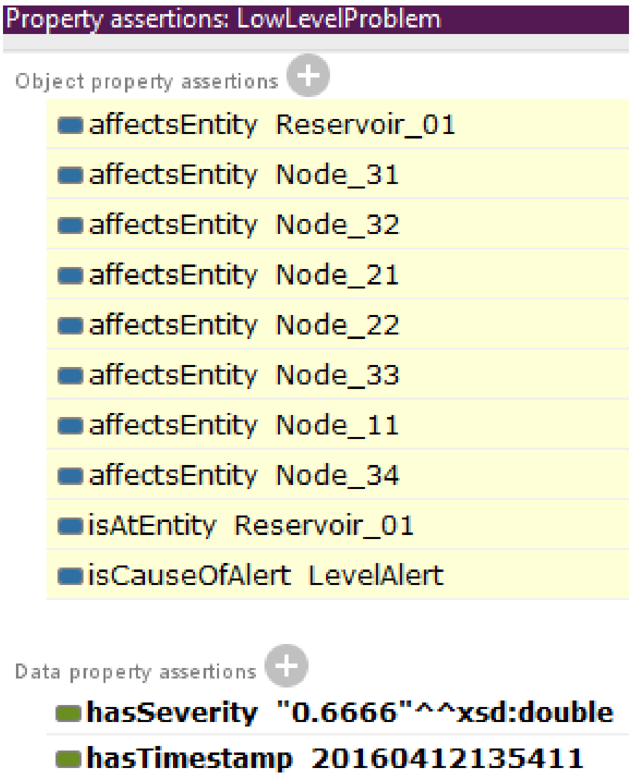


**Figure 66: Excerpt of resultant Abox knowledge after problem and alert inference testing**

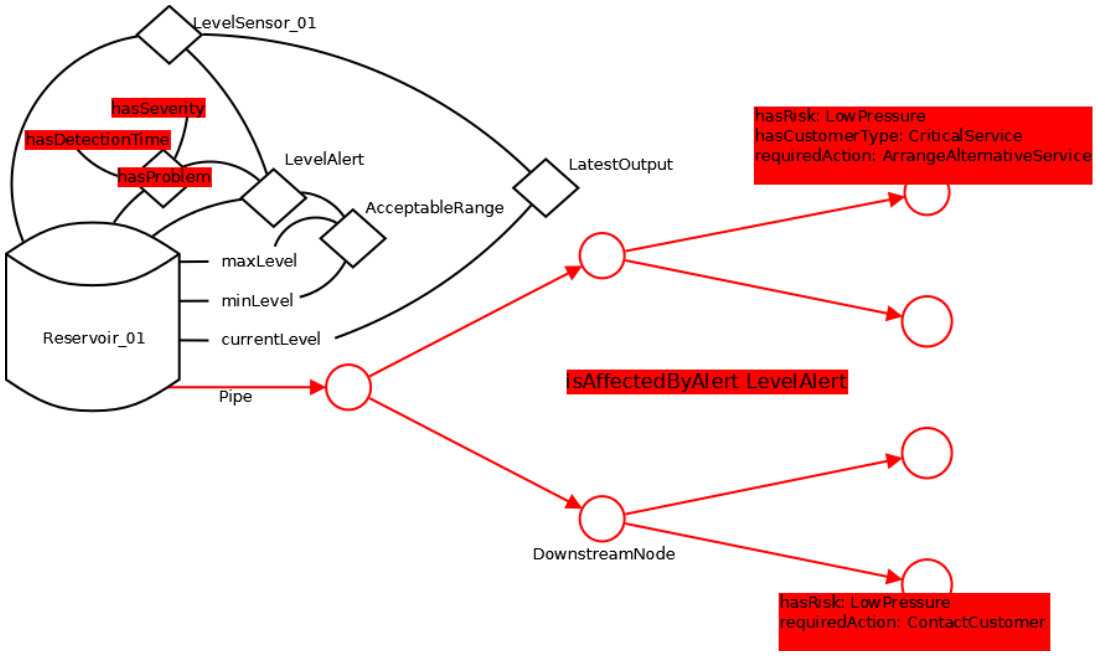


**Figure 67: Key knowledge inferred and extendable through the alert and problem inference testing**

#### 4.3.9.4 GRAPHICAL INTERFACE

A decision support tool was developed as a proof-of-concept, to demonstrate the benefits of the proposed platform, within the context of the established use case. The tool aimed to extend the state of the art of GIS tools, as well as typical water utility dashboards. The following section describes the interface from a user and technological perspective, and then provides evidence of the software's performance.

The GUI made several queries to the back end described: the first use of the knowledge management platform was a call to the Hypercat API to retrieve a list of pilot sites and knowledge bases, followed by a second call to discover the online sensors and alerts. The discovered SPARQL endpoints of each sites' knowledge base were then queried to retrieve descriptions of the water network objects such as pipes, assets, and sewer overflows. Once a sensor had been selected, as described in the following subsection, the timeseries endpoint was queried to retrieve its data for graphing. On discovering an active alert, the SPARQL endpoint was queried to retrieve further information such as the affected network entities and the problem's severity, based on the results of the inference engine's algorithms.

The decision support tool used the Google Maps API [572] to visualise the water network assets in an intuitive manner in a web-browser. The tool's functions were programmed in JavaScript and AJAX. CSS was used to mobile-optimise the page, such that it could be used on-site and away from utility workstations. Information about the assets was made available through context boxes on clicking assets. On clicking a sensor, an info window was displayed which showed a graph of the sensor's latest readings, which could be expanded for further investigation, as illustrated in Figure 68.

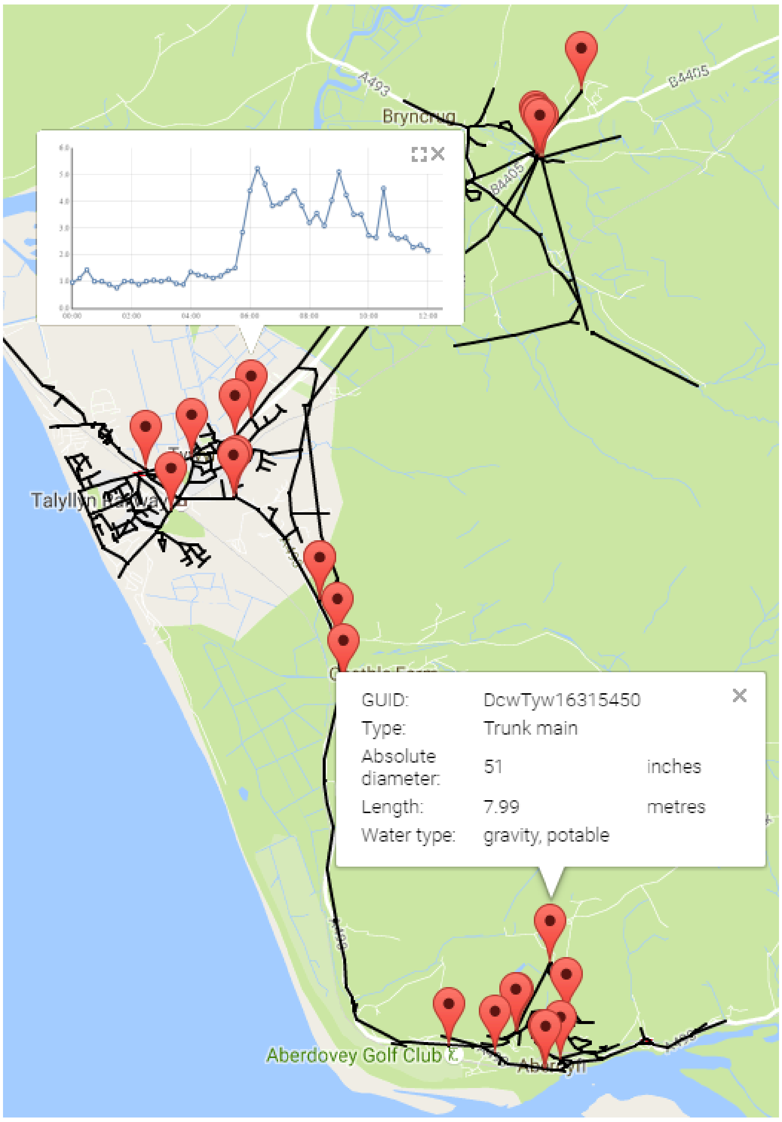


**Figure 68: Illustration of the user interface in network monitoring mode**

If a fault was detected by the inference engine, a side menu was uncovered and an alert icon was shown. On clicking this icon, the alert investigation state was entered, the colour of the affected network entities was changed to the alert's colour, and basic alert information was shown in the side menu. The user could then click the alert information to see more detail, or could click network entities to view information about them, as illustrated in Figure 69.

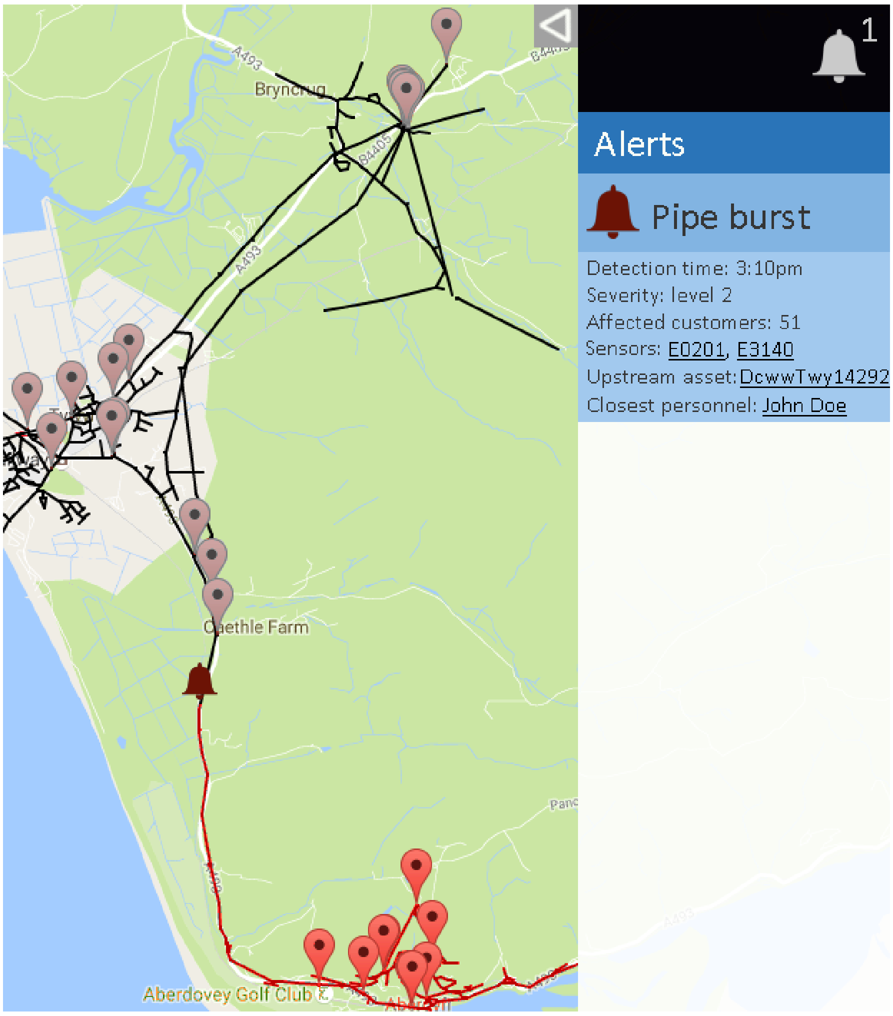


**Figure 69: Illustration of the user interface in alert mode**

### 4.3.9.5 SOFTWARE PERFORMANCE

The response time to network faults was evaluated `in vitro'. The total response time was evaluated as the sum of the time for the sensor reading to be updated in the knowledge base, the time taken for the inference engine to reason over the updated readings, and the time taken for the active alert to be represented in the GUI. The first of these three components depends entirely on the sensor network and communication technologies at a site, so was omitted.

The time taken for the inference engine to reason over the knowledge base was 450ms on average following caching. As this time is sub-second, it is deemed satisfactory. The time taken to retrieve knowledge about the active alerts from the SPARQL endpoint was found to be 550ms after caching, with an initial time of 3000ms, and with memory consumption increasing from 113MB to 800MB after caching. As sensor readings are only typically reported up to every 15 minutes in water networks, the overall latency observed was deemed satisfactory.

## 4.4 SEMANTIC WEB OF THINGS PLATFORM

### 4.4.1 OVERVIEW AND USE CASES

Following the conclusion of stage 2 through in-depth smart water domain action research, stage 3 then aimed to build on the work conducted to unify it towards broader smart city domain relevance. The aim of this work was to produce further evidence in consideration of the 3rd and 4th research questions, by bringing together the domain-specific learnings from stage 2 within a unifying software development and knowledge modelling design project, conducted independently. The goal was then to elicit deeper understanding of the causal relationships underlying the observations made throughout the 2nd stage, to clarify and evidence further the value of semantic technologies in smart cities, and to generalise the learnings towards relevance in other smart domains.

The primary use case of the project was smart city resource discovery and semantic integration, wherein a software developer or amateur programmer would be able to use the software to discover the web resources available to them and also contextualise the services offered by each resource. It was intended that the software could be deployed by a local authority to publicise their smart city capabilities and encourage development on their open data at both the grassroots and professional levels. The aim was to make a proof-of-concept of software which could be flexible enough to be used manually as a reference to assist a project's stakeholders, and also to act as a benchmark implementation of semantic integration and smart city visualisation for developers to build on top of.

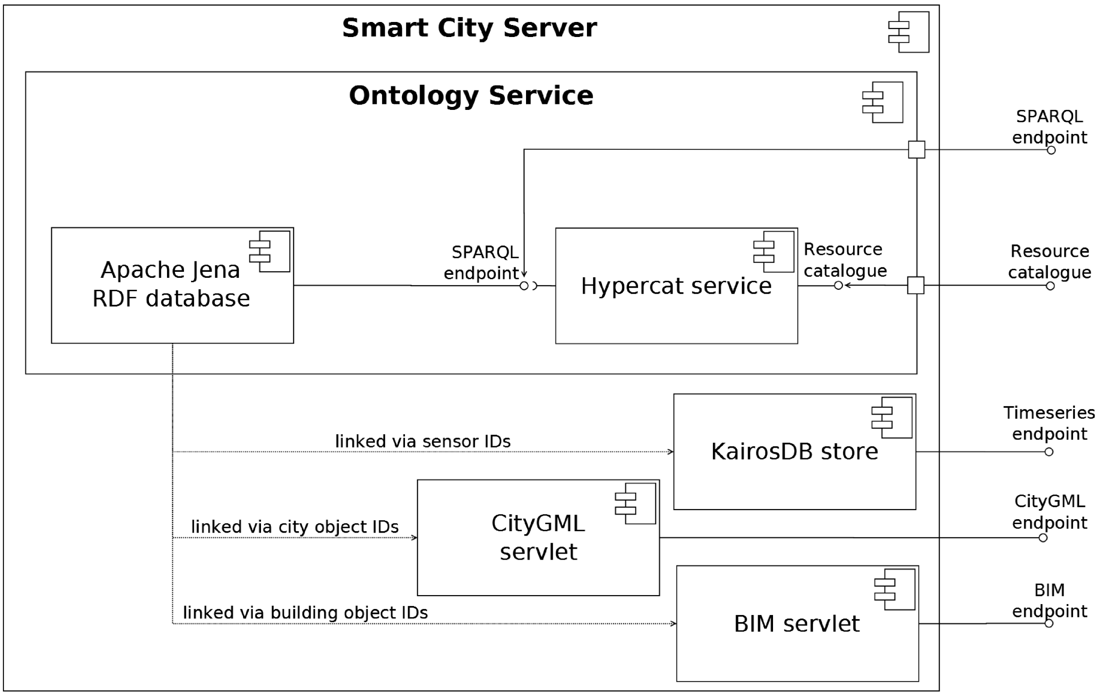


**Figure 70: Architecture of the smart city back-end software developed**

The project produced 3 primary components: a 3D graphical user interface, a ‘smart city server’, and a smart city ontology, as illustrated in Figure 70. The front end builds on the Cesium.js library [573] to show 3D city objects on a terrain model, with interactive functions to allow the user to view static data about objects, discover services and IoT Things in the area, view dynamic timeseries data from sensors, and view 3D BIM models of the buildings in the area. The smart city server integrated a number of modules into an Apache Jetty server to provide timeseries, BIM, CityGML, Hypercat, and SPARQL endpoints to the city’s data. The SPARQL endpoint allows the most functionality as it directly queries the Fuseki triple store which serves as the underpinning graph database for the software, whilst the other interfaces provide functions which are familiar and well-adopted within their smart city discipline. The triple store hosts an instance of the developed smart city ontology, which represents a conceptualisation of the smart city domain, aligned with several prominent ontologies and semantic resources of relevance to the target application.

This section now discusses each of these components in turn, beginning with the value proposition of the front end and its potential usage and performance. The smart city server is then presented, both in terms of how it supports the front-end, and how it serves as a reference implementation to be built on for domain-specific applications. Finally, the smart city ontology is presented and the results of its competency question testing are reported.

### 4.4.2 CUSP DEMONSTRATION USER INTERFACE

#### 4.4.2.1 OVERVIEW

The graphical user interface developed displays 3D city objects to the user as well as terrain and high-resolution imagery, and allows them to both interrogate the Things (sensors, other devices, and objects such as buildings) for data and resource discovery and also search the area for Things by text. The 3D city objects are initially rendered in a basic view to expedite loading, and reduce the mandatory data preparation requirements of the tool to promote widespread deployment. These objects can also be rendered using more detailed models on demand where they are available, or the user can move from ‘city mode’ to ‘building mode’ and interact with a building’s full BIM model where available. On selecting an object, data about the object is displayed in an information box. If the object has sensors deployed at it, these are linked to, and if the object offers services these are also linked. The user can then see data and metadata from sensors, and can interact with available services. The components which provide this functionality are now presented, and the uses and functions of the tool are described in more depth.

#### 4.4.2.2 SOFTWARE COMPONENTS

The web application is coded in a single HTML file and accompanying JavaScript file, with dependencies on a number of external libraries. The HTML document contains a single div element which contains the app. The HTML file is served from the root of the Jetty server so the app is the default landing page.

The application, written in JavaScript, is grounded in the Cesium.js library [573] for 3D geospatial visualisation, uses AJAX to query the smart city server, uses the xbim-viewer library [574] for the BIM model visualisation, uses the N3.js libraries [575] for handling client-side RDF data, and uses Chart.js [576] to display graphs from the timeseries data. The application itself then integrates the use of each of these libraries through a number of functions called either on loading the webpage, or on interacting with Cesium objects.

The application starts by creating and configuring the Cesium environment, and the query strings needed to retrieve data from the SPARQL and timeseries endpoints. A query is then made to the SPARQL endpoint to retrieve all of the basic city object geometries. If successful, the returned data is then parsed into an N3 graph and used to instantiate Cesium entities. Another query is then performed to instantiate entities for each sensor.

When a sensor is clicked, its description is shown in the information box and its dynamic data is presented in a graph. This is achieved by creating new div and canvas elements to hold the graph, then querying the timeseries endpoint to retrieve data from the sensor. If the query is successful a Chart.js line graph is instantiated on the canvas element to show a snapshot of the sensor's observations.

Finally, on clicking to move into 'building mode', the BIM file is retrieved from the server based on the building's ID in the triple store, and this is rendered in a translucent overlay div by using the xbim-viewer library. The xbim libraries offer mechanisms which could extend this functionality to suit target use cases.

#### 4.4.2.3 GUI FUNCTIONALITY

The various functions of the GUI are now presented and described. On loading the web app, the user is presented with the Cesium globe, zoomed to the basic object models loaded. For the demonstration instance developed, only buildings were instantiated, and these were arbitrarily coloured orange. The Cesium globe, the landing screen, and the basic city view are shown in Figure 71 - Figure 73. Loading the app took 6.36s (5.94s – 6.96s over 5 tests) on a local development version of the software without any optimisations. On average, 58% of this loading time was spent performing scripting activities, as shown in Figure 74. A significant part of this time was caused by displaying topography, and Cesium overheads, rather than the application's functions.

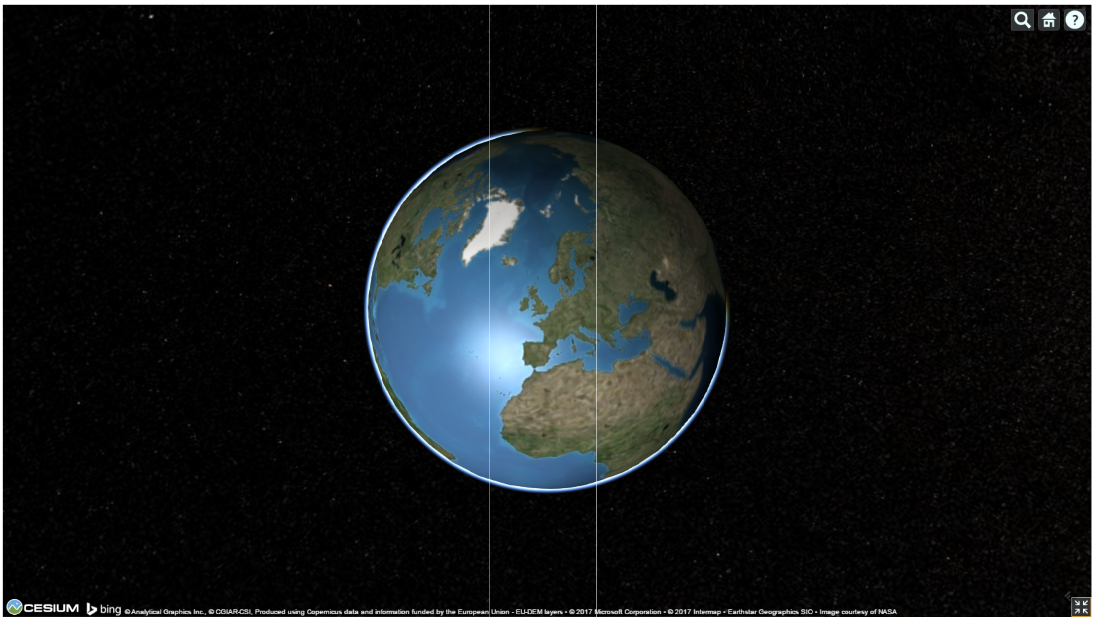


**Figure 71: Cesium globe in full screen mode**


**Figure 72: Default home screen showing basic entity geometries**

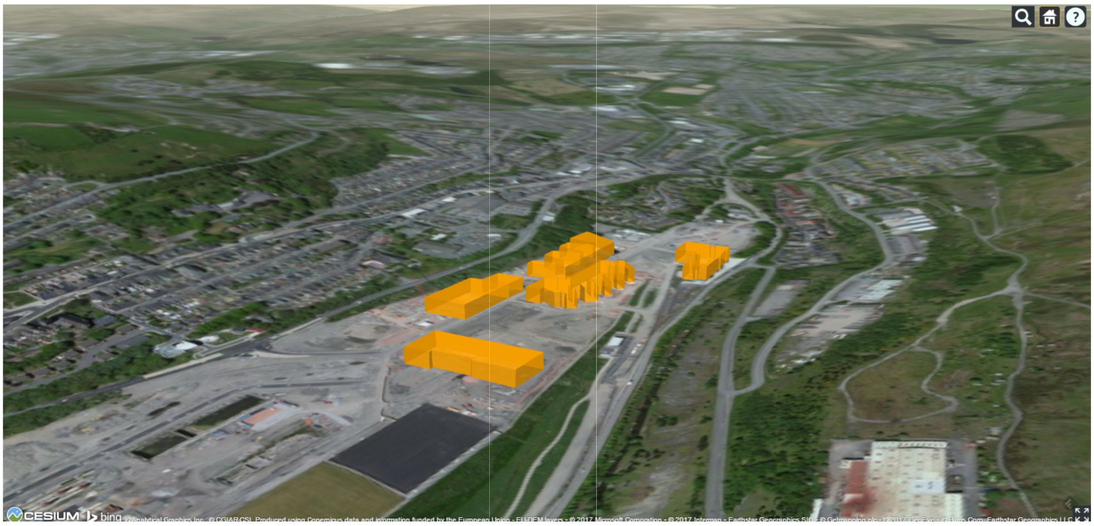

Figure 73: Screenshot of GUI showing 3D buildings and topography

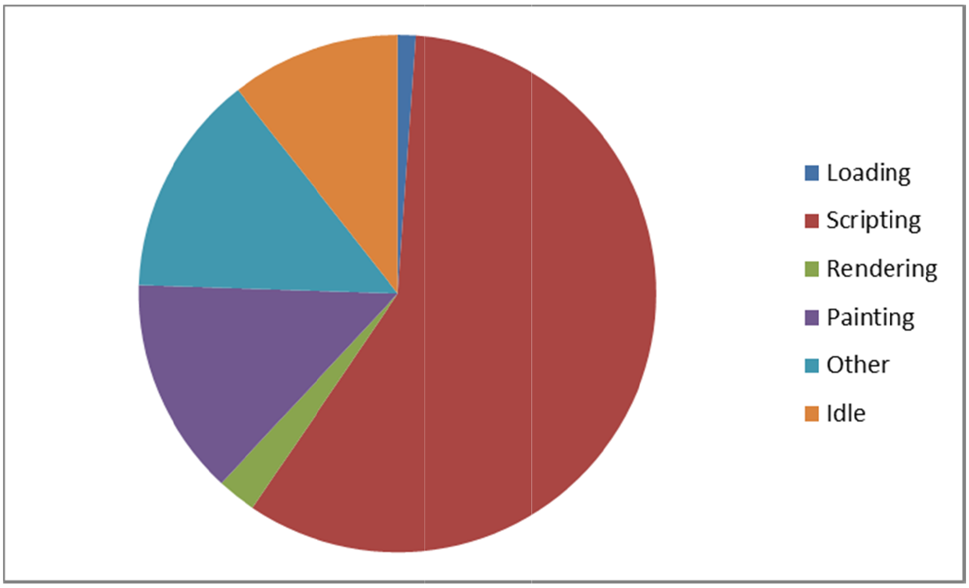


Figure 74: Breakdown of time spent on each activity whilst loading the application GUI

On clicking a building, the selection is indicated by an icon, and an information box is displayed, as shown in Figure 75. The information presented demonstrates possible uses of this function, but would be adapted to suit the intended use. There is no noticeable delay in loading this data. If a large number of city objects were present, this data could be queried on selecting an object to reduce the upfront load time.


**Figure 75: Screenshot of smart city GUI showing information box with example data**

On clicking a sensor's name, the information box changes to displaying information about the sensor, and a snapshot of the sensor's information is displayed in a new div, as shown in Figure 76. The user can also click the video camera icon in an object's information box to zoom to that object. A sensor's information box also contains a link back to the object it's deployed on. The time taken from the mouse up event over the sensor's name to the chart animation beginning is approximately 0.5 seconds with a local server deployment on a laptop, when retrieving 273 data time-value pairs.


**Figure 76: Screenshot of GUI with timeseries data snapshot and building focus shown**

On clicking the BIM link in a building's information box, the interface enters 'building mode', where the BIM model is visualised and users can interact with this, as shown in Figure 77. As the BIM model is stored in a format which is ideal for web display rather than raw IFC STEP format, loading, parsing and rendering a BIM model is very quick, taking approximately 1 second between the mouse up event and the end of the rendering process, for a detailed single storey building.

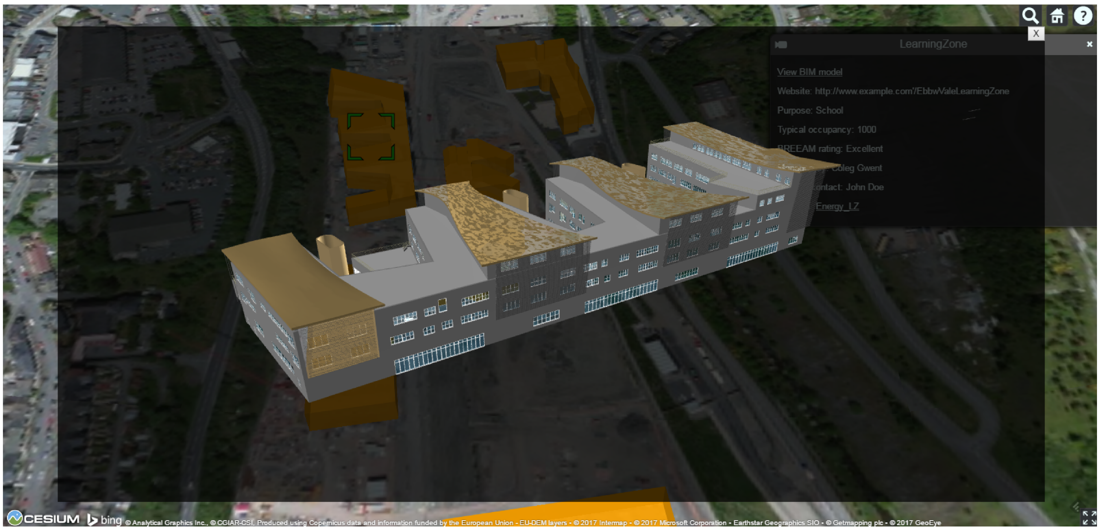


**Figure 77: Example building level interface showing BIM model overlay**

Also shown in the interface figures are icons in the navigation bar in the top right, which allow the user to search for places, objects, and things, as well as returning to the default location, and offering help on using the interface.

### 4.4.3 SMART CITY SERVER

#### 4.4.3.1 OVERVIEW

The 'smart city server' was developed to support knowledge management for built environment and IoT data, primarily as a proof-of-concept of the potential for semantic technologies to integrate and empower traditional approaches. The aim of this work was to act as a generic layer which could be built on for specific use cases or applications, as well as supporting the demonstration GUI developed. The server integrates several open-source libraries to provide interfaces to the smart city data which would be familiar to stakeholders from the various disciplines which constitute smart cities.

The smart city server was written in Java and makes use of the Apache Jetty project to expose the various libraries through web protocols. The main underpinning database is an Apache Jena triple store, which works alongside a KairosDB timeseries data store, as these are well suited to handle complex static data and large volumes of simple timeseries data respectively. As well as interfaces to these two components, the server provides interfaces to IFC files and CityGML files through the IFC toolbox [577] and CityGML4j toolbox respectively [578]. Finally, the server supports the discovery and interoperability of resources through a Hypercat interface, which uses a custom-made library to serve descriptions and 'API signposting' for the resources available in the area. This section now describes the API for these interfaces before presenting more implementation details about the components.

#### 4.4.3.2 API SPECIFICATION

The API for the smart city server is separated by component, as these are each accessed through separate paths, as shown in Table 31.

**Table 32: The URL paths of the components of the smart city server**

| Relative Path | Service | Description |
|---|---|---|
| **/** | GUI | for visualisation & manual discovery of resources |
| **/data** | KairosDB timeseries endpoint | Timeseries database for sensor data |
| **/sparql** | Fuseki SPARQL endpoint | Standard SPARQL endpoint for querying the knowledge base |
| **/bim** | BIM server | Custom interface for querying IFC resources through HTML binding of the apstex IFC Java Toolbox [577], as well as visualising BIM models. |
| **/citygml** | CityGML server | Custom interface for querying CityGML resources through HTML binding of the CityGML4j library [578]. |
| **/cat** | Hypercat endpoint | Endpoint for retrieving a catalogue of |

resources available through the server

The most powerful interfaces are the SPARQL and KairosDB endpoints, which provide access to the graph database and the timeseries data, including links to IFC, CityGML, and Hypercat objects. The IFC and CityGML files are stored as STEP and GML files, rather than triples, although this data could be moved to the triple store if relevant RDF formats are standardised. The SPARQL endpoint could then query all of the objects and properties on the server.

Objects, such as buildings, are identified by a URI, which can be resolved in a browser to an instance of the GUI, zoomed to that building with its infobox open. Building individuals have properties which link to their IfcBuilding and CityGML Building counterparts, where geometries are described. This is also the case between the ontology and sensor data in KairosDB. The SPARQL endpoint offers a standard API[579]; as specified in Appendix D.

Requests made to the timeseries interface are handled by a standard instance of KairosDB. The output from this service is a well-structured JSON object containing the requested time stamped data. An excerpt of the time series API is specified in Table 32.

**Table 33: POST method API for the KairosDB endpoint of the smart city server**

| POST /data | |
|---|---|
| Description | |
| A KairosDB query is passed to the endpoint in the body of the request as a JSON string specifying the ID of the sensor, as well as the desired date range, aggregation, time zone, grouping, return order, and maximum number of returned points. | |
| Response | |
| JSON | A JSON object detailing the number of data points returned, the sensor ID returned, other echoes of the query processed, and finally the data points. |

The Hypercat endpoint uses a novel Hypercat4j library and Hypercat Servlet as a mask for the SPARQL endpoint. This offers the API specified in Table 33 to Table 35.

**Table 34: API specification for the Hypercat root endpoint**

| GET | /cat | |
|---|---|---|
| Description | | |
| Top endpoint for retrieving a catalogue of available resources. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | Hypercat catalogue complying with the BSI:PAS 212 specification. | |

**Table 35: API specification for the Hypercat item endpoint of the smart city server**

| GET | /cat/{item_ID} | |
|---|---|---|
| Description | | |
| Retrieves information about a specific Hypercat item. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | Hypercat item description complying with the BSI:PAS 212 specification. | |

**Table 36: API specification for the Hypercat item description endpoint of the smart city server**

| GET | /cat/{item_ID}/description | |
|---|---|---|
| Description | | |
| Retrieves the name and a human-readable description of a Hypercat item. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | JSON object with two parameters: name and description, the values of which are Strings. | |

The BIM endpoint binds the BIM toolbox [577] to standard RESTful HTTP methods. This uses the IFC files stored on the server. Requests are handled by the developed BIM Servlet. On each function call, the ID of the IFC file is used to load the file, before performing the required methods on it. This approach uses less memory than loading all of the models on start-up, but alternative approaches such as storing the IFC data in a database exist. A proof-of concept of the binding was developed, which offers the API specified in Appendix D, and partially in Table 36.

**Table 37: Top level endpoint of the BIM interface of the smart city server**

| GET /bim | | |
|---|---|---|
| Description | | |
| Provides a name, description, and size of all the BIM models stored in the server. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | Object containing an array of objects, each containing three parameters: name, description, and size. | |

The CityGML endpoint, in parallel to the BIM endpoint, binds the CityGML4j library to RESTful HTTP methods. Again, this interacts with GML files at present but could interact with CityGML data in the triple store or another database. Requests are handled in a similar manner to the BIM interface, in that the CityGML file is loaded, parsed, and queried for each call. Again, the work conducted produced a proof-of-concept of this binding, offering the API specified in Appendix D and summarised in Table 37 and Table 38.

**Table 38: Top level endpoint of the CityGML interface of the smart city server**

| GET /citygml | | |
|---|---|---|
| Description | | |
| Top level endpoint; returns the names, descriptions, and sizes of all the CityGML files stored on the server, where available. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | Description | |

**Table 39: Entity level endpoint of the CityGML interface of the smart city server**

| GET /citygml/{id}/{gmlid} | | |
|---|---|---|
| Description | | |
| Returns the properties of a CityGML element | | |
| Request parameters | | |
| id | String | ID of the citygml model, used to find the correct file |
| gmlid | String | ID of the element to be retrieved |
| Response | | |
| JSON | JSON object with a key:value pair for each property returned. | |

Finally, the Jersey Servlet developed to handle GUI requests offers the functionality of linking directly to Smart City objects through the API specified in Table 39 and Table 40.

**Table 40: Root endpoint of the smart city server, for the GUI**

| GET / | | |
|---|---|---|
| Description | | |
| Root of the smart city server; returns the default GUI | | |
| Request parameters | | |
| None | | |
| Response | | |
| HTML | The GUI JavaScript dynamically creates the HTML file | |

**Table 41: Object level endpoint of the GUI for resolving object URIs to human-readable information**

<table>
<tr><th colspan="3">GET /{id}</th></tr>
<tr><td colspan="3" align="center">Description</td></tr>
<tr><td colspan="3">City object URL, deferences to an instance of the GUI zoomed into that object with its properties visible</td></tr>
<tr><td colspan="3" align="center">Request parameters</td></tr>
<tr><td>id</td><td>String</td><td>ID of the entity to be dereferenced</td></tr>
<tr><td colspan="3" align="center">Response</td></tr>
<tr><td>HTML</td><td colspan="2">The ID is passed to the GUI JavaScript, which creates the default instance and then performs a Cesium flyTo function to zoom onto the target object.</td></tr>
</table>

#### 4.4.3.3 HYPERCAT RESOURCE DISCOVERY SERVICE

As an emerging standard, Hypercat has less tooling already available than the other components of the server. Therefore, a new Java library was developed, entitled Hypercat4j. This represents Hypercat items and catalogues as Java objects and provides functions for interacting with these, including the retrieval of items and catalogues in JSON.

The Item class has two member variables: ‘String href’ and ‘SetMultimap<String, String> metadata’ (using SetMultimap from com.google.common.collect). The href is the URL of the item, as specified by PAS212 [47], and the metadata object is a set of key:value pairs, where each key can have multiple values, but no duplicate key:value pairs can exist. Item objects are instantiated from an href and a description, which are both mandatory according to PAS212. A human readable description is added to the metadata object with the key “urn:X-hypercat:rels:hasDescription:en”. Member functions expose an item’s data according to the API specified in Table 41 and Table 42.

**Table 42: Javadoc specification of the Item class of the developed Hypercat4j library**

| Return Type | Method and Description |
|---|---|
| void | **addMetadata**(String _rel, String _val)<br>Adds the provided rel:val pair to the item’s metadata, if the pair is not already present. |
| String | **asJson**()<br>Returns the item’s serialisation as a string in JSON format complying with BSI:PAS212 |
| String | **description**()<br>Gets the item's human readable description |
| void | **description**(String _desc)<br>Updates the item’s description |
| String | **href**()<br>Gets the href of an item, which is the url that the item is available at |
| void | **href**(String _href)<br>Updates the item’s href |
| SetMultimap<String, String> | **metadata**()<br>Gets all of the item's metadata as a SetMultimap |
| String | **metadataAsJson**()<br>Returns the item’s metadata array, serialised as a string in JSON format |
| void | **updateMetadata**(String _rel, String _val)<br>Removes any previous values in pairs where the provided _rel is a key, then adds the provided rel:val pair |
| Set<String > | **values**(String _rel)<br>Returns a set of all of the values associated with the provided rel in the item’s metadata. Returns null if not found. |
| String | **valuesAsJson**(String _rel)<br>Returns all of the values associated with the provided rel, serialised as a JSON array string. Returns null if not found. |

The catalogue class extends the item class to also include a ‘Set<item> items’ member variable, which represents a collection of Hypercat items. The Catalogue constructor also instantiates a catalogue from an href and a description, but adds an additional metadata key:value pair specifying that the object stores Hypercat catalogue content. The use of Sets to store items and metadata ensures that each catalogue can't have duplicate items. The catalogue class's explicit methods are specified in Table 42.

**Table 43: Javadoc specification of the Catalogue class of the developed Hypercat4j library**

| Return Type | Method and Description |
|---|---|
| void | `addCat(java.lang.String _href,String _description)` Adds a catalogue to the catalogue as an item by calling the catalogue constructor with the provided href and description |
| void | `addItem(Item _item)` Adds an item to the catalogue by directly adding an item object to the member set of items |
| void | `addItem(String _href,String _description)` Adds a new item to the catalogue by calling the item constructor with the provided href and description |
| String | `asJson()` Returns the entire catalogue as a JSON object string, based on the serialisation specified in BSI:PAS212 |
| Item | `item(String _href)` Returns an item object with the specified href if present, otherwise returns null. Note that each href must be unique within the scope of each catalogue. |
| String | `itemAsJson(String _href)` Returns a JSON object string of the item with the specified href, or returns null if not found. |
| Set<Item> | `items()` Returns a set of all the item objects in the catalogue. |
| String | `itemsAsJson()` Returns a JSON array string of all the items in the catalogue, without the catalogue's metadata or href. |

On receiving a request for a Hypercat catalogue or an item description, the Hypercat Servlet produces a SPARQL query based on the requested information and sends this to the triple store. The retrieved information is then parsed and converted into Hypercat objects and ultimately a JSON string describing the specified Things, which is then returned to the client.

#### 4.4.3.4 PERFORMANCE TESTING

The smart city server was tested for performance on a Samsung 900X laptop with 8GB of RAM and an Intel i5 processor, using the demonstration instance displayed in Figure 72 - Figure 76. The server was restarted 10 times and took an average of

10.7s +/- 0.15s to start up, and circa 260MB of RAM whilst idle, and occupied 570MB of hard drive space. This includes 230MB of git files for version management, 215MB of triple store data, and Java source files for the smart city server.

On the first load of the GUI, the memory demand of the server increased by 2.8 MB, after spiking by 20.2 MB, based averages of 6 tests. On further requests for the GUI, familiar caching behaviour was observed.

To gauge the scalability of the approach, the query time was tested for exponentially larger numbers of buildings until failure, whilst restarting the server between tests to avoid caching. The query response times from this testing are shown in Figure 78 below on a logarithmic scale. Following testing with a million buildings, the response time was regarded as a failure and no further testing was conducted. The relationship of response time (t) to number of buildings (n) was linear, with a function of $t = 0.1n + 757$ and an $R^2$ value of 0.9998. The maximum amount of memory used by the experiment was circa 2GB.

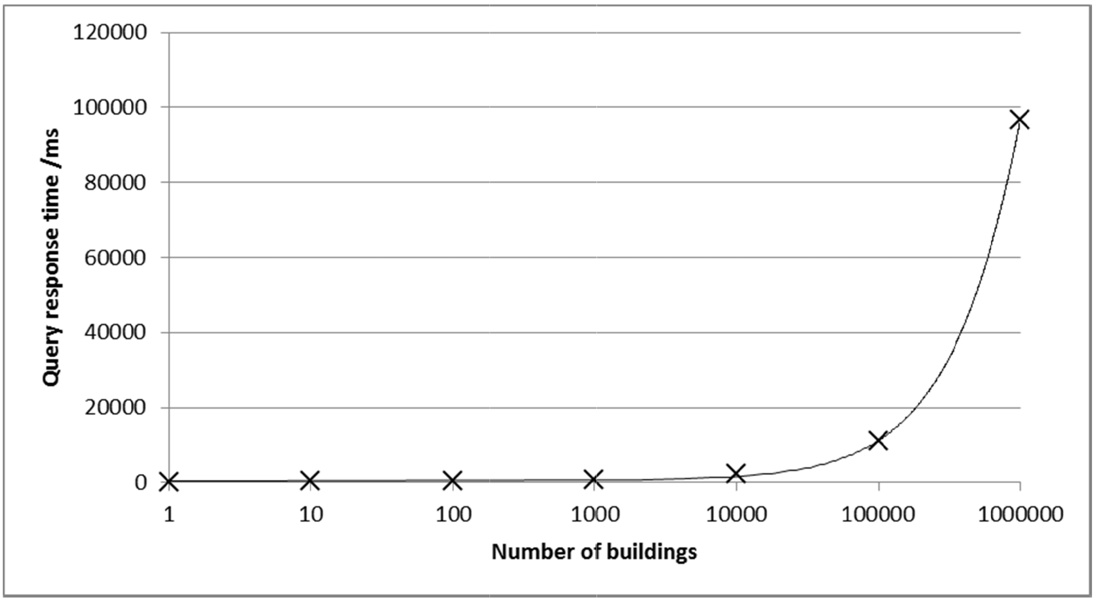


**Figure 78: Scalability testing of smart city server**

### 4.4.4 SMART CITY SEMANTIC MODEL

The smart city server used a graph database to integrate knowledge between its separate components and interfaces. This was based on a comprehensive smart city upper ontology, which described the concepts and relationships in the domain, as an alignment and extension of various existing models and standards pertinent to smart cities. This section describes the outputs of the ontology curation process and competency question testing.

#### 4.4.4.1 SCOPING AND REQUIREMENTS SPECIFICATION

The ontology development process was similar to that of the smart water ontology. The outputs of the earlier stages of this process were a set of initial statements informally bounding the scope of the ontology, a set of candidate donor ontologies, use cases, and finally a set of competency questions.

There were two main use cases for the ontology: contextualising discovered resources, and integration of data and semantics across smart domains. The former was derived from the main use case of the overall smart city platform as described previously. The latter use case emerged from the learnings from the energy and water domains. Based on this 'initial intent', a conceptualisation of the scope of the ontology emerged, which was captured through informal scoping statements. These included that the ontology should:

- Describe systems and networks in general for extensions into smart domains
- Capture the symmetry between social, physical, and cyber systems
- Include modelling patterns across IoT Things (physical objects and devices), their web representations, and the data and services they expose
- Include modelling patterns across agents, the services they offer, the web representations of these, and the places they are offered from when not via web protocols

A set of candidate donor semantic models were then established for potential reuse. These were selected based on the degree of consensus and adoption they had achieved, and their relevance, as described in Table 43.

Several models were not available in OWL format. For the BSI:PAS182 smart city concept model, an OWL model was developed based on BSI:PAS 182. For the Industry Foundation Classes, the latest ifcOWL conversion was used. For CityGML, the semi-automated conversion from the University of Geneva [580] was used. For Hypercat, a conceptual model was derived from the BSI:PAS212 specification and formalised in OWL. Whilst several IoT ontologies were considered for reuse, none emerged as a de facto standard, and there was little value in adopting a dedicated IoT ontology as well as the Hypercat, SAREF, and SSN ontologies already identified.

**Table 44: Models reused in the smart city semantic model**

| Acronym/name | Description | Owner | # Entities | Date |
|---|---|---|---|---|
| **Smart City Concept Model** | Lightweight generic smart city ontology including organisations, services, places, and events. | BSI | 54 | 2014 |
| **IFC** | Open format for the exchange of building information models. | buildingSMART | 768 | 2016 |
| **CityGML** | Information model and XML encoding for representation, storage, and exchange of 3D city and landscape models. | OGC | 557 | 2012 |
| **Hypercat** | Lightweight JSON-based hypermedia catalogue format for exposing collections of URIs. | BSI | 2 | 2016 |
| **SAREF** | Describes smart appliances and their services, functions, properties, and commands. | ETSI | 223 | 2013 |
| **QUDT** | Schema for describing quantities, units, dimensions, and types. | QUDT | 229 | 2017 |
| **SSN** | Describes sensors, observations, and related concepts. | W3C | 107 | 2011 |

Given the informal scope produced, competency questions were formalised to guide the ontology development, such as:

- What devices are present?
    - What services do these devices expose over the web, and what are their endpoints?
- What sensors are present, what do they observe, and how can their observations be accessed?
- What is the context of the data produced by sensor X?
    - Where is the sensor deployed?

    - What property does it observe, and what is the metadata of these observations?
- What agents are present and what services do they offer?

#### 4.4.4.2 ALIGNMENT AND REUSE OF EXISTING MODELS

Following the scoping stage, the identified models were merged and pruned into an initial ontology, to be further extended. This was a manual and judgement-based process, given the available definitions of concepts in the donor ontologies. The Dolce Ultra Lite ontology was used to provide abstraction to the donor ontologies. Where concepts were to be reused, new concepts were formalised, and an owl:equivalentClass or owl:equivalentProperty relationship was made. This resulted in an ontology which partially met the specified scope, depending on the depth of the donor ontologies. For example, the IFC and CityGML models contain detailed taxonomies of physical urban objects, whereas none of the ontologies represented web services in much depth. Several concepts were in the scope of multiple ontologies, so complex semantics emerged around heterogeneities.

Regarding the superclass 'Agent', most of the models described some aspect(s) of agency, especially with regards to human agency, and so bringing these varying perspectives into the coherent structure shown in Figure 79 was challenging. Smart City Ontology (SCO) is the target ontology, and the other prefixes are defined in Table 44.

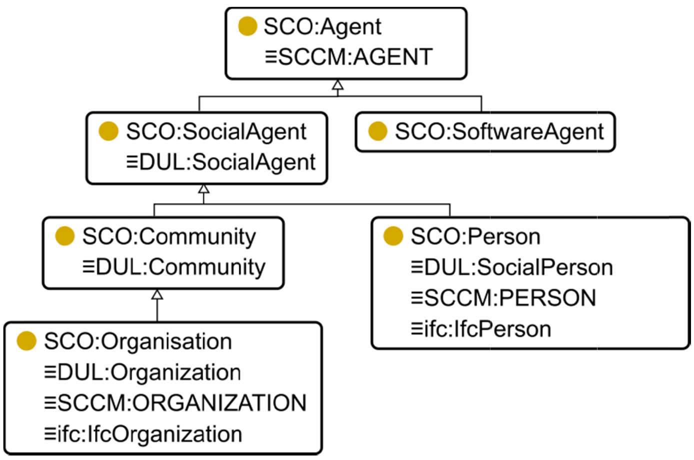


**Figure 79: Excerpt of SCO agent taxonomy and equivalencies**

Table 45: Prefixes used in the context of the smart city ontology

| Prefix | Meaning | Full URL |
|---|---|---|
| SCCM | Smart City Concept Model, BSI:PAS 182 | http://www.semanticweb.org/SCCM# |
| DUL | Dolce Ultra Lite | http://www.ontologydesignpatterns.org/ont/dul/DUL.owl |
| SSN | Semantic Sensor Network Ontology | http://purl.oclc.org/NET/ssnx/ssn# |
| HC | Hypercat | N/A |
| QUDT | Quantities, Units, Dimensions, Time | http://qudt.org/schema/qudt/ |
| SAREF | Smart Appliance Reference Ontology | http://ontology.tno.nl/saref# |
| CGML | CityGML | http://www.opengis.net/citygml/<module_name>/2.0/ |
| IFC | Industry Foundation Classes | http://ifcowl.openbimstandards.org/IFC4_ADD2# |

Regarding physical city objects, CityGML adopts strong semantics from a location-oriented world view, whereas DUL, SSN, SAREF, and SCCM adopt weaker semantics, and so a complex hierarchy emerged in the target ontology, as shown in Figure 80.

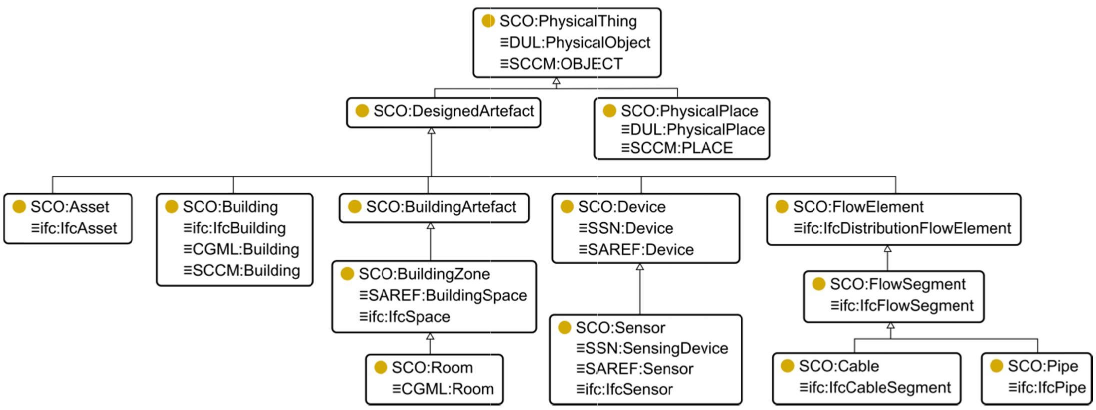


Figure 80: Excerpt of SCO physical thing taxonomy and equivalencies

Regarding social entities, DUL describes social objects as concepts which exist within a communication event, which leads to a complex pattern of modelling physical and social aspects of humans separately. This is somewhat unique to DUL, and was challenging to align with the other donor ontologies' perspective of humans as indivisible. Ultimately this led to the conceptually awkward notion of an observation being a social entity, but the overall resulting taxonomy was coherent, as shown in Figure 81.

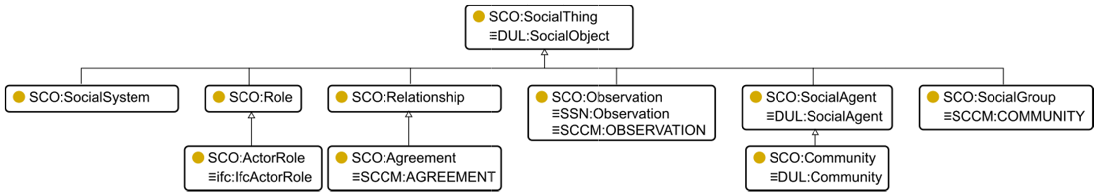


**Figure 81: Excerpt of SCO social thing taxonomy and equivalencies**

A number of other important equivalencies were stated following the initial reuse process, some of which are presented in Table 45. Overall 121 explicit equivalencies were formalised, which does not include many which may be inferred by a semantic reasoner.

**Table 46: Excerpt of miscellaneous pertinent SCO class equivalencies**

| Relative IRI in Target Ontology | Equivalent Concept(s) in Donor Ontology |
|---|---|
| **WebService** | SAREF:Service |
| **Service** | SCCM:SERVICE |
| **Event** | DUL:Event, SCCM:EVENT |
| **Action** | DUL:Action |
| **Task** | SAREF:Task, ifc:IfcTask |
| **InformationThing** | DUL:InformationEntity |
| **Quantity** | QUDT:Quantity |
| **Record** | SCCM:ACCOUNT |
| **TimeSeries** | ifc:IfcTimeSeries |
| **TechnologicalSystem** | ifc:IfcSystem |
| **Quality** | DUL:Quality |
| **Property** | SSN:Property, QUDT:QuantityKind, SAREF:Property |

#### 4.4.4.3 SMART CITY ONTOLOGY

The smart city ontology extended the reused concepts in both depth and breadth. Breadth was primarily added by modelling web concepts to offer better 'API signposting' for discovered web resources, as well as system theory and topology concepts. This section briefly outlines the main class hierarchies and modelling patterns developed.

The core focus of the ontology was contextualising the data and web services offered by a smart city, so the SSN ontology was reused extensively in developing the modelling patterns in this area, resulting in the pattern presented in Figure 82.

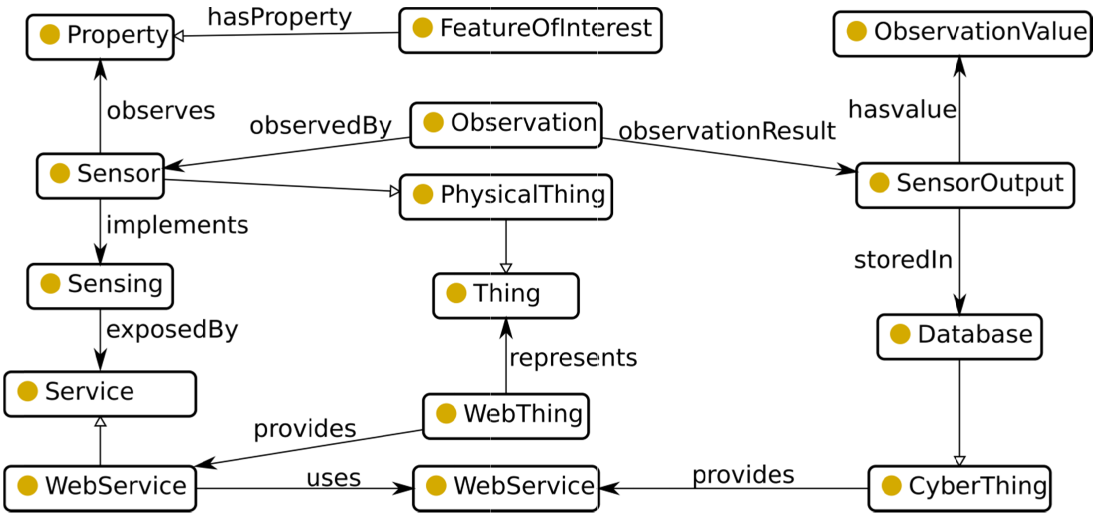


**Figure 82: Modelling pattern for exposing sensor observations via web things**

The topology of physical city objects resulted in considerable depth of modelling, as shown by Figure 83.


:PhysicalThing ≡ wgs84_pos:SpatialThing ≡ DUL:PhysicalObject ≡ SCCM:OBJECT
:Commodity ≡ saref:Commodity
:Electricity ≡ saref:Electricity
:Gas ≡ saref:Gas
:Water ≡ saref:Water
:DesignedArtefact
:Asset ≡ ifc:IfcAsset
:PowerAsset
:Substation
:WaterAsset
:PumpingStation
:Bridge ≡ brid:Bridge
:Building ≡ SCCM:BUILDING ≡ bldg:Building ≡ ifc:IfcBuilding
:BuildingArtefact
:BuildingElement ≡ ifc:IfcBuildingElement
:BuildingFurniture ≡ ifc:IfcFurniture ≡ bldg:BuildingFurniture
:BuildingStorey ≡ ifc:IfcBuildingStorey
:BuildingZone ≡ saref:BuildingSpace ≡ ifc:IfcSpace
:Room ≡ bldg:Room
:Door ≡ bldg:Door ≡ ifc:IfcDoor
:Railing ≡ ifc:IfcRailing
:Ramp ≡ ifc:IfcRamp
:Roof ≡ ifc:IfcRoof
:ShadingDevice ≡ ifc:IfcShadingDevice
:StairFlight ≡ ifc:IfcStairFlight
:Wall ≡ ifc:IfcWall
:Window ≡ bldg:Window ≡ ifc:IfcWindow
:CityFurniture ≡ frn:CityFurniture
:Device ≡ ssn:Device ≡ saref:Device
:Actuator ≡ ifc:IfcActuator
:Alarm ≡ ifc:IfcAlarm
:Appliance
:AVAppliance ≡ ifc:IfcAudioVisualAppliance
:LightFixture ≡ ifc:IfcLightFixture
:EnergyConversionDevice ≡ ifc:IfcEnergyConversionDevice
:Boiler ≡ ifc:IfcBoiler
:Chiller ≡ ifc:IfcChiller
:Generator ≡ ifc:IfcElectricGenerator
:HeatExchanger ≡ ifc:IfcHeatExchanger
:PVPanel ≡ ifc:IfcSolarDevice
:Transformer ≡ ifc:IfcTransformer
:Sensor ≡ ifc:IfcSensor ≡ ssn:SensingDevice ≡ saref:Sensor
:FlowMeter ≡ ifc:IfcFlowMeter
:Valve ≡ ifc:IfcValve
:FlowElement ≡ ifc:IfcDistributionFlowElement
:Fitting ≡ ifc:IfcFlowFitting
:Junction
:FlowSegment ≡ ifc:IfcFlowSegment
:Cable ≡ ifc:IfcCableSegment
:Pipe ≡ ifc:IfcPipeSegment
:FlowSource
:FlowTerminal ≡ ifc:IfcFlowTerminal
:Pump ≡ ifc:IfcPump
:StorageElement ≡ ifc:IfcFlowStorageDevice
:ElectricityStorageElement ≡ ifc:IfcElectricFlowStorageDevice
:Tank ≡ ifc:IfcTank
:Valve ≡ ifc:IfcValve
:TechnologicalSystem ≡ ifc:IfcSystem
:TransportationObject ≡ tran:_TransportationObject
:AuxiliaryTrafficArea ≡ tran:AuxiliaryTrafficArea
:Path
:Railway ≡ tran:Railway
:Road ≡ tran:Road
:Square ≡ tran:Square
:Track ≡ tran:Track
:TrafficArea ≡ tran:TrafficArea
:Tunnel ≡ tun:Tunnel
:PhysicalGroup
:PhysicalPlace ≡ DUL:PhysicalPlace ≡ SCCM:PLACE
:Platform ≡ ssn:Platform
:VegetationObject ≡ veg:_VegetationObject
:Bush
:PlantCover ≡ veg:PlantCover
:Tree
:WaterBody ≡ wtr:WaterBody

**Figure 83: Physical Thing taxonomy for the smart city ontology**

The relationship of physical devices to their related web services was also central to the scope, and is illustrated in Figure 84. Conceptual modelling was conducted to allow the description of basic APIs, to allow simple API signposting where public data is distributed via GET requests to an open endpoint. More complex RESTful APIs can be captured in a dedicated format through a RAML document, which can directly link to a web thing or any of its components.

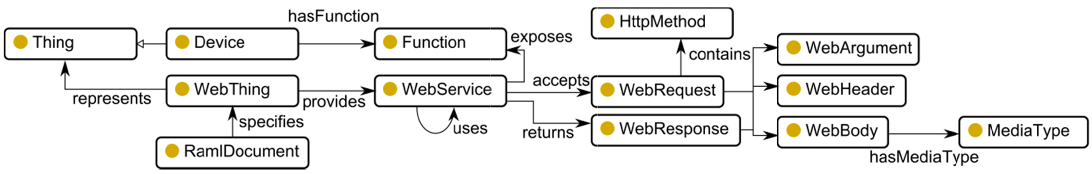


**Figure 84: Web Thing and API modelling pattern**

The relationship between physical and social entities and their abstraction to system theories and topologies was very relevant, and can be described using the pattern shown in Figure 85.

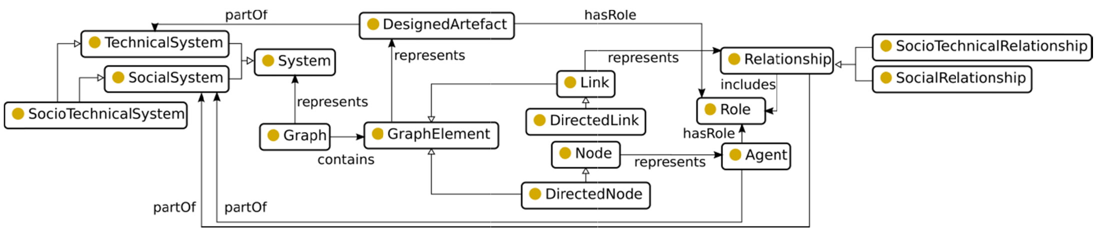


**Figure 85: Socio-technical system modelling pattern for smart city ontology**

The nature of services as either web-based or real-world was also relevant, as highlighted by the BSI:PAS 182 smart city concept model, and this was represented using the pattern shown in Figure 86. This also shows an object property between an agent and a designed artefact: the SCO facilitates two levels of detail in modelling relationships. Specifically, a social or socio-technical relationship can be expressed as a named individual, if the application needs to refer to aspects of the relationship, such as the details of a contract, and/or the relationship can be expressed through a single object property. Note that ‘hasAgencyOver’ is a super object property, whose sub properties include ‘owns’ ‘manages’, ‘controls’ etc.

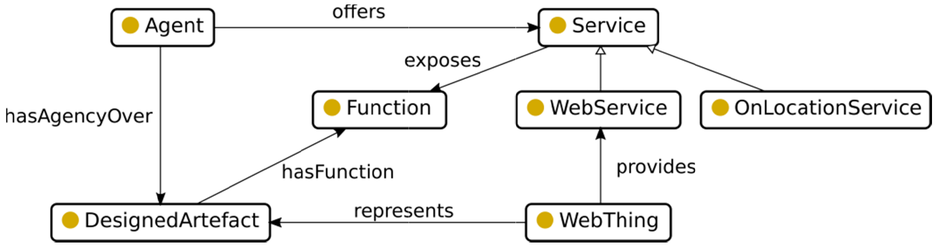


**Figure 86: Overview of the SCO service modelling pattern**

The developed Smart City Ontology represents a direct extension of the BSI:PAS182 Smart City Concept Model (SCCM), as introduced by Figure 87, which shows an excerpt of the alignments between the SCO and the SCCM, as well as the extensions. The main added breadth is the modelling of sensor networks, socio-technical systems, and IoT Things.

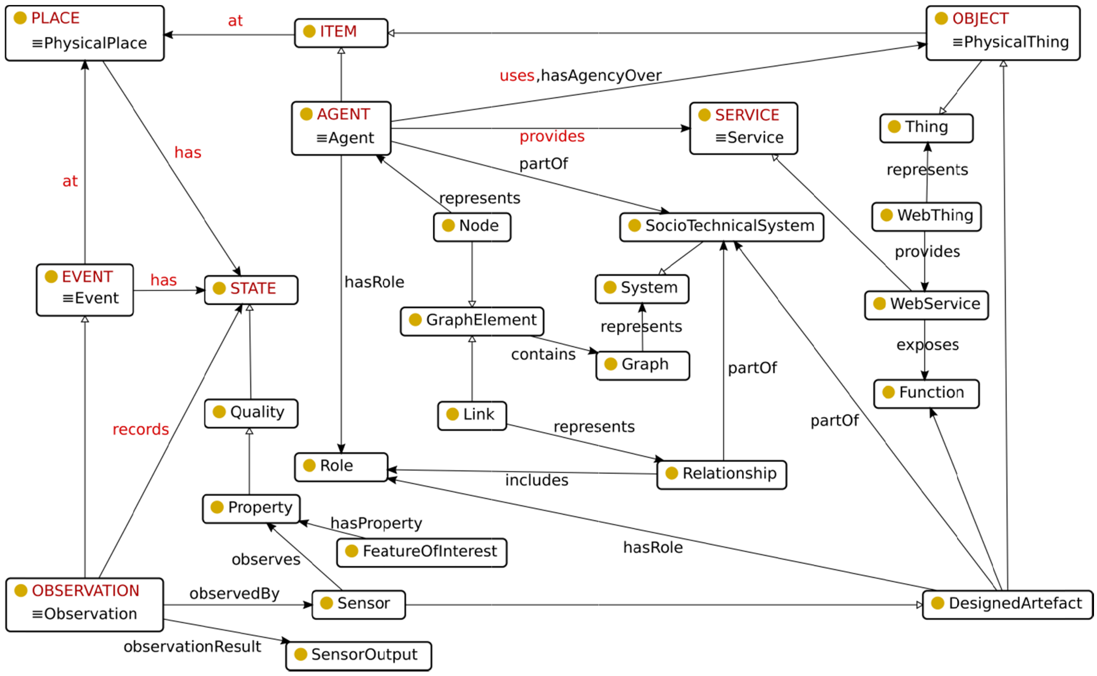


**Figure 87: Excerpt of ontology to illustrate extensions of SCCM (red: SCCM, black: SCO)**

#### 4.4.4.4 EXTENSION TO SMART DOMAINS

The second key use case targeted in the development of the Smart City Ontology was promoting interoperability and alignment between smart domains. Following the action research undertaken in stage 2 in the energy and water domains, the perspectives adopted in those domains were aligned through the Smart City

Ontology. This primarily made use of the system modelling and physical modelling concepts in the SCO, but also included some cyber and social concepts. This illustrates that the Smart City Ontology is intended as an upper ontology for smart cities, and can be applied within smart domains by creating taxonomies underneath the provided classes. This is introduced in Figure 88 with simple taxonomies for energy and water.

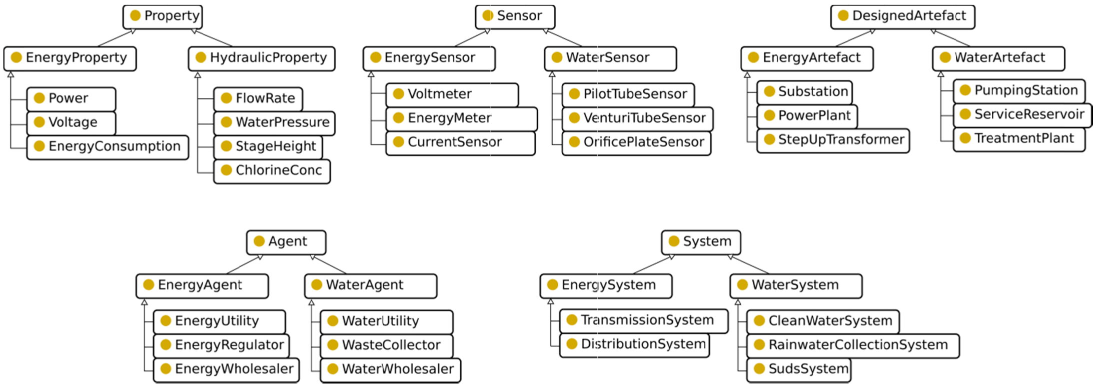


**Figure 88: Illustration of the SCO as an upper ontology with domain extensions**

#### 4.4.4.5 VERIFICATION AND COMPETENCY QUESTION TESTING

Verification and validation of an ontology is essential in assuring its quality and usefulness as a candidate representation towards domain consensus. This quality assurance is necessarily an iterative process, whereby the initial scope is used to guide the development of the ontology through to the maintenance stage of its lifecycle, whilst the scope and requirement specification should also be revisited in light of any challenges, breakthroughs, or new domain knowledge acquired during development. Once at a maintenance stage, the ontology is deemed as acceptable by its modellers, but should be revisited and refined regularly to reflect any changes in expert perspectives of the domain. Verification is the first stage of this quality assurance process, and assures that the ontology uses valid syntax and logic, such that no contradictions are present and the ontology represents *a* world view. The next stage, validation, aims to assure that the world view described by the ontology is an accurate and sufficient formalisation of the domain bounded by the scope and competency questions. The first stage, verification, was conducted automatically, using the Protégé software tool, and also using the OntOlogy Pitfall Scanner (OOPS) developed by the Ontology Engineering Group of the Technical University of Madrid [581]. Both tools verified that the ontology is valid: Protégé successfully reasoned over the ontology with the HermiT reasoner, and no inconsistent logic was identified. The OOPS tools declared that the ontology had passed all ‘pitfall’ tests apart from 1 minor issue where two class identifiers were deemed to be synonyms in the SSN ontology, and several minor issues where classes were ‘unconnected’: they had no explicit properties linking them to other classes.

After verification, a preliminary validation was conducted using the prescribed competency questions, formalised as SPARQL queries. These tests were passed, an excerpt of which is illustrated in Table 46 and Table 47.

**Table 47: Sensor discovery example competency question testing evidence for smart city ontology**

| Natural language question: | |
|---|---|
| What web enabled sensors are there? | |
| SPARQL query | |
| PREFIX sco:<http://www.semanticweb.org/sco#><br>PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#><br><br>SELECT ?sensor ?webThing<br>WHERE {<br>?sensor rdf:type sco:Sensor .<br>?webThing sco:represents ?sensor .<br>?webThing rdf:type sco:WebThing .<br>} | |
| Output (csv format) | |
| Sensor | webThing |
| sco:sensor_01 | sco:webThing_sensor_01 |

**Table 48: Sensor context example competency question testing evidence for smart city ontology**

| Natural language question: |
|---|
| What is the context of the data produced by sensor X? |

SPARQL query

```
PREFIX sco:<http://www.semanticweb.org/sco#>

PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>


SELECT ?sensor ?platform ?property ?unit ?lat ?lon

WHERE {

?sensor rdf:type sco:Sensor .
?property rdf:type sco:Property .
?observation rdf:type sco:observation .
?sensorOutput rdf:type sco:sensorOutput .
?observationValue rdf:type sco:observationValue .
?unit rdf:type sco:unit .
?location rdf:type sco:location .
?sensor sco:observes ?property .
?sensor sco:deployedOn ?platform .
?platform sco:hasProperty ?property .
?sensor sco:hasLocation ?location .
?location sco:hasLat ?lat .
?location sco:hasLon ?lon .
?observation sco:observedBy ?sensor .
?observation sco:observationResult ?sensorOutput .
?sensorOutput sco:hasValue ?observationValue .
?observationValue sco:unit ?unit .
}
```

Output (csv format)

| sensor | platform | property | unit | lat | lon |
|---|---|---|---|---|---|
| sco:sensor_01 | sco:building_01 | sco:enConsump_01 | sco:kilowatt-hours | 11.01 | 12.24 |

## 4.5 SUMMARY OF RESULTS

This chapter has presented the outputs and results of the 3-stage investigation. Firstly, the results of the theoretical study were presents, which built on the literature review and engagement with experts to propose a system of systems conceptual framework for smart cities and a model which integrated the emerging IoT stack with BI theory. Next, the use cases pursued to evaluate the role of semantic technologies were specified for the individual energy and water domains, and finally at the energy-water nexus.

Next, the outputs of the participatory action research iterations were presented, first in the energy sector, followed by the extended learning iteration around a smart water platform. For both energy learning iterations, an overview of the project, software outputs, ontologies, and testing results, were described. In the water sector investigation, significantly more work was conducted, which resulted in various requirements engineering outputs, a detailed ontology, several instantiations of this, a semantic rule engine and suite of rules, and finally a proof-of-concept decision support tool.

Finally, the outputs of the 3rd stage of the investigation were presented; a smart city semantic web of things platform and upper level smart city ontology. Each of the components of this investigation were described in turn, beginning with the user interface and front-end value proposition. Next, the back-end software was described and the results of its performance testing was offered. Finally, the smart city semantic model was described, including its own scoping artefacts, modelling patterns, alignments, and testing results.

The following discussion chapter critically considers the evidence presented in this chapter towards answering the research questions. This begins by firs considering each action research iteration in turn, before then unifying these in a discussion of the overall contributions to the literature, and finally discussing the relevance to practice.

# 5 DISCUSSION

This chapter presents a discussion of the work conducted and outputs presented in the previous chapter. The analysis of the results relies primarily on qualitative data, as is typical of action research investigations [538]. The data most pertinent to the hypothesis are the qualities of the designed artefacts, the perspectives of the experts engaged with, and the learnings of the researcher. As mentioned in section 3.5.2; 4 aspects are required to validate design research outcomes: artefact success, generalisation, novelty, and explanation capability [560]. Each of these 4 aspects can be argued through discussion of the aforementioned data. By discussing the value which semantic technologies brought to each of the systems and experts engaged with, knowledge regarding the role of semantic technologies in future system designs and problem solutions emerges.

The chapter is structured into 3 sections; firstly, direct analyses of the action and design research conducted in the 2nd and 3rd stages of the methodology are offered, proposing an interpretation of the results, the lessons learned and challenges faced through the process. Secondly, a discussion is made of the overall work conducted as a contribution to the discourse within various academic fields. Finally, the relevance of the work to industry is discussed, and recommendations are made for practitioners.

## 5.1 ANALYSIS OF PROJECT RESULTS

Three iterations of action research undertaken in the 2nd stage of the methodology are reported on here, coinciding with the 3 main research projects engaged with through the course of the investigation, followed by a discussion of the 3rd stage of the methodology at the end of this section. Firstly, the work conducted in the building energy domain is discussed, before the work conducted in the energy domain at the multi-consumer and grid levels, and finally the work in the water domain. Within each action research iteration, each section firstly discusses the work conducted and the outputs produced in a general sense, before specifically discussing the use of semantic technologies. The discussion is summarised in Table 48, which highlights the value observed of the semantic approach in each action research iteration.

**Table 49: Overview of the observed benefits of the semantic approach in each action research iteration**

| **Building energy management (KnoHolEM)** | |
|---|---|
| Traditional Approach | Semantic Approach |
| Data mining would require ad-hoc elicitation of data semantics. Making the simulation and mapping the simulation rules to real-time data would require another investigation of the installed sensors, and analysis of CAD data. The rule engine would need another ad-hoc mapping to the SQL database with non-standard queries. | Semantic knowledge base integrated the fuzzy reasoner, data mining rules, federated CAD data, and simulation-based optimised rules. Use of semantic web standards allowed the GUI to be thin. Adding further modules would be simplified. |

| **Smart grid demand side management (MAS2TERING)** | |
|---|---|
| Traditional Approach | Semantic Approach |
| Agent message payloads would have to be carefully considered to ensure each agent correctly understood its meaning. Developing the communication framework would require a more complex process of analysing the purpose of each message type and its required sentiment. Integrating the MAS with web services would require ad-hoc mapping of the agents' world views to the data used in the web services, which must be duplicated for each new web service. | Agents communicated through a shared language and understanding of data because of the JavaBeans ontology. Developing the communication framework for each scenario could be systematically aligned with the message payloads and their meanings. The MAS was integrated with web services through the OWL version of the ontology. Further web services or agents could be added with significantly less effort. |

| Smart water management (WISDOM) | |
|---|---|
| Traditional Approach | Semantic Approach |
| All mappings between the system components' semantics would need to be ad-hoc and manual. Discovery of available sensors would also be manual, or would require another ad-hoc solution. Integration with legacy systems would require a detailed analysis of legacy semantics for each component. The applications would not have a coherent view of the water network and available resources. Developed applications and services would be much 'thicker' as they would have to manage significant domain complexities. Building new applications on the platform would require more effort to find and understand the available data and services for each new application. | The various applications access data through a coherent data model with clear semantics, masking the heterogeneity of the legacy systems. The intelligent core services consider the domain at a conceptual level without a need to manage conflicting data semantics. The proof of concept GUI was thin and simple to make. Data can be incorporated from other systems with less effort, and from instances of the aligned semantic models with very little effort. Further applications or core services could be added without a detailed analysis of domain semantics. The ontology can be reused in other systems, which could then integrate with the WISDOM system. |

## 5.1.1 SMART BUILDING ENERGY MANAGEMENT: KNOHOLEM

### 5.1.1.1 CRITICAL ANALYSIS OF ACTION RESEARCH

The retrofit building energy management system developed combined the use of theoretical and empirical approaches to optimal rule generation, and also incorporated negotiation between the FM and the optimisation engine. Negotiation is achieved by varying the termination goal of the multi-objective optimisation process in an iterative manner. Through interaction with the GUI, the FM may experiment with the desired energy reduction in order to gauge its effect on the key performance indicators within the facility such as PMV, temperature or luminance. This recognises the importance of including an FM within the decision making process. Therefore, the system does not aim to replace the FM as the decision maker, but aims to better inform the FM.

The developed retrofit BEMS demonstrated its capability of delivering energy savings through analytics across existing data sources and actuators in a building, by using semantic middleware to integrate heterogeneous devices within a cloud based, service-oriented architecture. As well as the novelty of the semantic approach, the solution represents a step change by encouraging the use of AI by FMs, by respecting the FM's role in the decision process and by using an engaging GUI, and the solution has been successfully deployed in a public building in The Netherlands.

The developed tool goes beyond existing solutions, as they typically only serve data to monitoring tools, without providing actionable insight or higher order knowledge, which the proposed system accomplishes. This served to better empower the decision maker in making more informed choices. Valuable knowledge was produced from the data through its contextualisation in the triple store, and the integrated use of artificial intelligence and optimization algorithms. These applied business objectives of reducing energy consumption and improving occupant comfort, within a set of constraints and through a set of decision variables, to the incoming data. Finally, offering this intelligence in real time was accomplished by the fuzzy reasoner, which used SWRL rules to decide suggested actions. This was an iterative negotiation process with the human expert, who interrogated the suggestions offered by the solution and utilised expert knowledge to alter the prescribed objectives and constraints, then finally accept and actioned an acceptable solution.

The system saved the expert time and offered insight into the optimal setpoints to choose based on predicted weather patterns and the building environment. In the testing of this system, a high degree of trust was established between the FM and the proposed actions, although the FM did have the option at each iteration of not trusting the suggestion and simply changing the decision space until a suggestion was presented which was deemed trustworthy. Further work could automate the process completely, although this may raise regulatory issues in many situations.

One significant outcome of the case study was that the use of simulated data successfully supplemented legacy sensor installations, significantly mitigating hardware investment costs. The case study also highlighted the importance considering legacy system integration into any developed solution. This was evident in the pilot building where sensors were closely coupled with existing legacy BEMS systems, and so retrieving data from these systems was a prerequisite to building the system.

#### 5.1.1.2 USE OF SEMANTIC TECHNOLOGIES

The case study considered a whole BEMS solution through its development and testing from the perspective of the hypothesis being tested, and this section discusses the solution's use of semantics to integrate data and ICT resources. The system relied on a knowledge base in order to integrate the fuzzy reasoner, data mining rules, federated CAD data, and simulation-based optimised rules, for use by the GUI. This knowledge base was produced for each pilot building as an instantiation of the OWL domain ontology developed by Krahtova et al. [568] within the case study. The knowledge base also included SWRL rules produced by the data mining process and simulation-based optimisation process. In this manner, the knowledge base for each pilot site stored the semantics required to apply intelligence and context to the data, to proceed from 'information' to an actionable request, in a coherent and shared manner. Further applications could be integrated into the system with relatively little effort.

It was found that the solution was required to include a traditional 'dashboard' interface, as well as the advanced analytics, for familiarity to the expert users. The users also requested the ability to interrogate the higher order knowledge was presented, by justifying it through its lower order constituents and the additional semantics and logic which led to the higher order knowledge. This was achieved through graphs of historical data for each sensor, and a traffic light system indicating the current state. Further work could present more detail of the logical argument which led the system to make each recommendation.

The performance of the semantic software was sufficient for the intended use cases. With the final number of rules and individuals, the system was able to return suggestions rapidly. However, the number of rules was significantly limited to ensure this response speed. Initially, the simulation-based optimisation process produced thousands of rules, but suitable performance was only achieved with circa 500 rules embedded in the knowledge base. If a large number of rules must be incorporated into a future system, it is suggested that these are handled by a dedicated application. Also, managing the triple store in a distributed manner or using a more scalable triple store could be explored.

The knowledge base stores all the data about the building and its systems relevant to the BEMS. The knowledge base integrates heterogeneous data sources and also provides intelligence capabilities through reasoning over the rules and structures contained in the knowledge base. Real-time sensor data was not stored in the Fuseki server, as this reduced the performance below acceptable levels, instead it was stored in an SQL-database, and was referenced through an ID in the semantic model.

### 5.1.2 SMART GRID DEMAND SIDE MANAGEMENT: MASTERING

#### 5.1.2.1 CRITICAL ANALYSIS OF ACTION RESEARCH

The action research conducted in the smart energy grid domain produced evidence pertinent to evaluating the hypothesis, across the research questions. The main work conducted was the development of an ontology for use in a multi agent system for demand side management. The action research iteration also contributed to and analysed the integration of this ontology in a knowledge management solution and the wider ICT solution. The primary use case of the semantic modelling conducted was prosumer demand side management; specifically with a close coupling of home appliance automation and the peak shaving goals of the smart grid. This was achieved through the concept of a holonic multi-agent system, and its emergent properties. Deep interoperability between the agents was essential for them to effectively engage in market-based negotiations and optimisation.

The holonic systems approach was utilised to promote optimal demand profiles, through the incentive of intelligent flexibility trading, whilst respecting individual desires and beliefs. The knowledge management architecture developed allowed the integration of both traditional agents, and web services, which didn't exhibit autonomy, but offered the advantages of modern web technologies. The ability of semantic technologies to facilitate a shared domain conceptualization and hence interoperability amongst virtual artefacts represents a critical enabling step towards highly distributed energy systems. The diversity and prevalence of interoperating components is increasing and is leveraged as an opportunity in holonic systems, but the ability to share and fully utilize data between these components is critical to their intelligent management.

In this case study, intelligent distributed management was achieved to improve the demand profiles of aggregated dwellings, in a consistent and reliable manner, whilst providing economic benefits to the active prosumers and meeting their preferences. The adaptability of the agent-based approach also promoted resilience in the system by optimising the emergent properties of the aggregated dwellings in the case of grid failure. Given that these benefits relied directly on pervasive interoperability, this highlights the value of the semantic approach adopted herein, in facilitating this interoperability in a powerful, confident, and open manner, as discussed below.

#### 5.1.2.2 USE OF SEMANTIC TECHNOLOGIES

The described ICT system relied on semantics to integrate the intelligent components on 2 levels; firstly, the ontology's axioms were formalised in JavaBeans to be utilised directly by the MAS, as a common language for message payloads. Secondly, the ontology, formalised in OWL constructs, was used to integrate the agents with web services such as load and RES predictions. This common language allowed confident and powerful interoperability with less effort than a manual, implicit approach, given the number of coordinating software components. The language also served to provide partial alignment between several neighbouring models, including CIM, IEC 61968-9, energy@home, IFC and SAREF.

It is worth noting that the JavaBeans model and the OWL model formalised exactly the same domain conceptualisation, as the JavaBeans model was created automatically from the OWL ontology, so they were semantically homogeneous. Most *in vivo* systems would probably contain several varying models, so incorporating this complexity into the system is a possible avenue of further work. The integration of the agents with web services was achieved through a semantic web approach which utilizes a domain ontology, within a triple store. This data can then be accessed through an ontology service wrapper around a SPARQL endpoint, based on the JENA ARQ API. The entities within the domain ontology also then comprise the content ontology for FIPA-ACL message payloads.

The developed domain ontology was based primarily on an OWL representation of reused semantic resources; the CIM, IEC 61968, OpenADR, and the energy@home data model. Firstly, classes and slots were elicited from the described resources, to produce a coherent model across the domains of smart appliances, demand response and smart grid, whilst still respecting the scope of the ontology prescribed by the use cases. The ontology also formalized the concept of domestic load flexibility, and included concepts related to the trading and aggregation of this flexibility. Given the difference in nature between the data schemas of the reused standards and that of OWL and JADE ontologies, the federation and re-use approach adopted represents a best-case for future compliance with existing standards if they are expressed normatively in an ontological format in the future.

### 5.1.3 SMART WATER: WISDOM

#### 5.1.3.1 CRITICAL ANALYSIS OF ACTION RESEARCH

This final iteration of action research within the second stage of the methodology aimed to explore further the evidence and artefacts produced previously, but within the water domain. This involved a significantly greater engagement with a research and development project leading a requirement engineering process and subsequently an ontology and software development process. The target system aims to deliver intelligent water sensing, analytics, services and interfaces, towards optimization of the water network at the utility level, as well as in homes, through interoperability and demand side management. A key innovation of the proposed solution, and the focus of the work conducted, is the integration of heterogeneous data sources and varied analytics and visualization components, through a domain ontology, which has been instantiated and deployed within a dedicated web service. The utility of such an approach has been highlighted within the network optimization service, as the ontology web service allows this to utilize data from across the water value chain at runtime.

The results presented show that the ontology and its software deployment are sufficient as a conceptualization of the water domain for use within a near real-time decision support system. The validation of the domain ontology displays that it is agreeable amongst a wide range of stakeholders within the industry, and that it could contribute significantly to the standards identified by as critical in the smart water domain [111]. The software testing conducted indicates that the performance of the ontology service and SPARQL endpoint, an extension of the Apache Jena APIs, was sufficient for the velocity and volume of requests and updates deemed typical within the target software platform's use cases.

One of the key analytics services provided by the WISDOM platform is the optimization of pump and reservoir management schemes in water networks. This uses a range of metaheuristic optimization techniques to minimize the energy and water consumption of the network by providing online near-optimal suggestions for pump and valve control. The ontology service plays a key role in facilitating this optimization of the water network, by integrating data across domains and scales for use by the optimization module such as network asset descriptions and CSO overflow locations.

An example of the integration of data across water systems is illustrated in Figure 89, which also demonstrates the key approach of ensuring data privacy; whilst the use of domestic knowledge is useful for analytics, and network knowledge may be helpful to consumers, it is critical to respect the privacy of data owners. The system therefore balances the benefit of integrating data with the requirement for data security and privacy by facilitating private and shared objects.

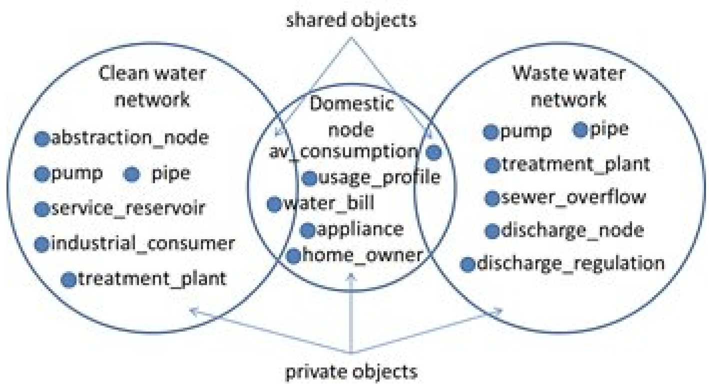


**Figure 89: Integration of object knowledge across the water value chain to highlight the capability for data privacy**

A use case which highlights the interoperability benefits of the semantic alignment at the building scale is shown in Figure 90, which illustrates the hypothetical case of a consumer with both a water feedback app and an appliance scheduling app interacting with their devices.

Figure 90 shows the objects (physical and otherwise) which are relevant to the repositories of both applications, including those which can be reused from WISDOM by simply aligning with SAREF. If a third application is introduced, the previous mappings to SAREF have already been completed, meaning that only one mapping is required to integrate the application, as opposed to mapping to both of the other applications. This is illustrated further in Figure 91 as a means to avoid exponential mapping tasks in the likely future case of many integrated software artefacts. Also, it is not required for one single common model to gain universal acceptance for the premise of Figure 91 to hold; even with 2 or 3 common models (each mapped to each other) the mapping task growth is mitigated significantly.

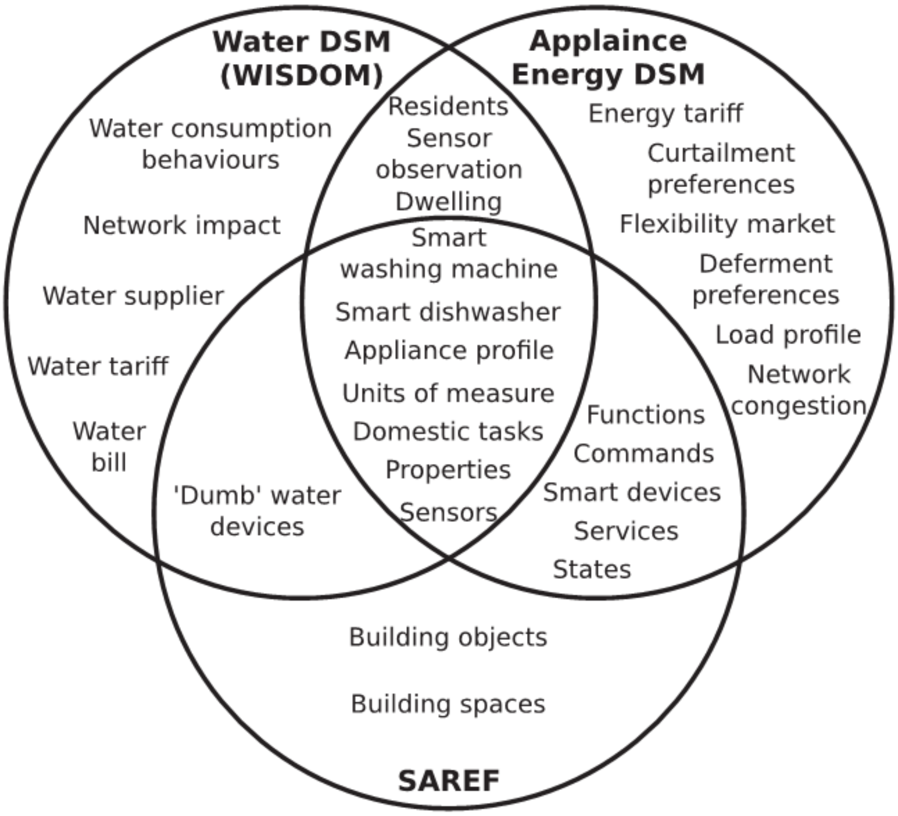


**Figure 90: Object reuse across smart home applications, through alignment with the SAREF ontology**

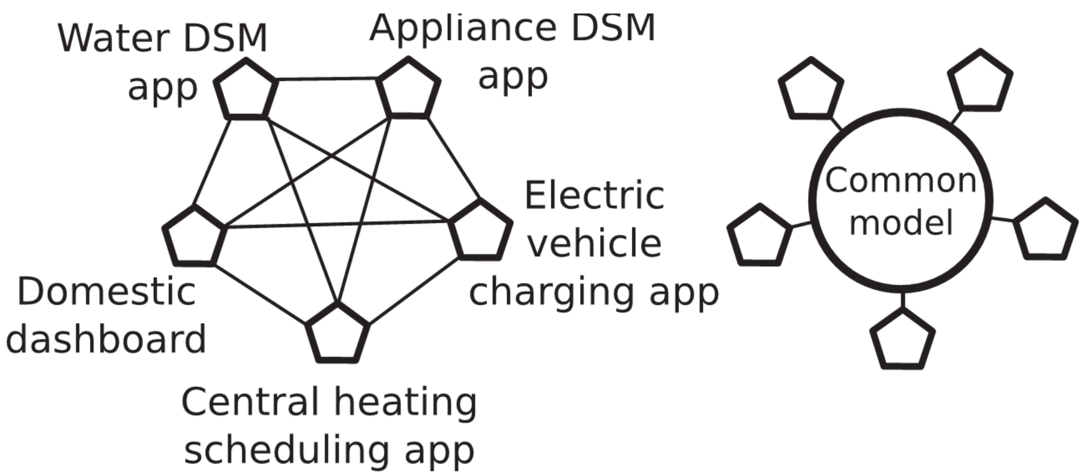


**Figure 91: Mitigation of mapping task growth with increasing entities through a common model**

#### 5.1.3.2 ROLE OF SEMANTICS IN TARGET ICT SOLUTION

A key aspect of the WISDOM project was the integration of data, analytics, and decision support components across the water value chain. This interoperation presents a significant challenge for existing technologies which use proprietary protocols, and convey messages with widely different terminologies and meanings.

Further to the benefits of semantic models in the domain in general, the WISDOM semantic models have a specific role within the target project and target software platform. They aim to capture sufficient water management knowledge for the implementation of the WISDOM scenarios through the WISDOM platform's business services, whilst also being suitable for reuse in other smart water systems. The WISDOM semantic models underpin the ICT platform by formalizing a vocabulary of technological, sensory and socio-economic concepts and their relationships within the water management domain. This provides a common interface for the software components to share data through, enriches sensed data with contextual meaning and utilizes inference to produce new knowledge from that which is explicitly inputted or reused from GIS data sources.

The approach prioritized requirements engineering, leading to an acknowledgement of the ontology's scope boundaries in a formal and rigorous way. This facilitated the reuse of the ontology in future applications, as its role alongside other ontologies becomes clearer. For example, it could be aligned with a model of treatment plant concepts to enrich and integrate data between high level system management and asset level performance objectives.

The ontology was built on previous efforts such as controlled vocabularies [504], [582], a format for hydrologic time-series data [508], and a data model of geospatial utility networks [512]. The main advantages of the ontology compared to alternative semantic resources are:

- The use of greater semantic expressivity (more detailed and machine-interpretable descriptions of meaning).
- Modelling the domain with more breadth and depth (including things like sensors, people and GIS descriptions).
- Using techniques to allow alignment with other smart city systems (e.g. smart grids) so that interoperability can be achieved with them in the future.

The deployment of the ontology as a web service supports the benefits of a service-oriented architecture [15] and hence allows plug-and play capability with other software components of the WISDOM architecture, and potentially beyond. The software was developed and tested on a local machine and has since been deployed on the secure cloud environment provided by Imperial College London for the purposes of the project. This software is now at a mature stage where it is able to handle real-time SPARQL requests as well as custom functions for the most common foreseen uses of the ontology service.

Regarding the semantic inference capabilities of the system, Apache Jena natively supports 4 types of reasoner based on the architecture: transitive reasoner, RDFS rule reasoner, OWL/Lite reasoners, and generic rule reasoner [554]. These increase in inference capability from the transitive reasoner to the generic rule reasoner, although even the most capable of the Jena reasoners typically achieves less inference than the Pellet reasoner, due to Jena being RDF based and Pellet considering the entire conjunctive query [583].

As the Jena API integrates the Pellet inference engine [583] with little effort, and the Pellet engine is well regarded for capability and speed, the Pellet reasoning engine was initially chosen to exploit the maximum potential from the OWL axioms. Further, Pellet relaxes OWL-DL restrictions on the OWL-Full features, and allows the majority of SWRL built-in atoms. This meant that Pellet could reason over rules which included maths features and numerical comparisons, which have been included in the developed SWRL rule set. However, during testing this inference capability was found to be slower, and less reliable, than the use of a separate Drools engine. For this reason, and given that a Drools engine matches the other benefits of Pellet, a Drools engine was ultimately chosen.

By performing semantic inference at the platform layer, the resultant knowledge could be shared by multiple applications, for example if the same issue affected different business processes and hence different expert decision makers. By offering a single point of truth with a range of data and contextualisation to various applications, the response of the organisation can be streamlined, coordinated, and more effective.

The rules developed to supplement the ontology’s semantic inference would assist with impact mitigation after detecting a network fault, by identifying the customers likely to be affected by the issue and exposing their details in a secure manner to the appropriate member of staff. Also, by including knowledge regarding the organisation’s performance indicators, the inference engine could offer targeted information and suggested actions. This would support the stage of resolution, which could also be supported by extending the knowledge-base to cover the asset management processes, people, and organisations, as this could identify nearby people able to resolve the issue.

### 5.1.4 SMART CITY SEMANTIC PLATFORM

The third and final stage of the methodology aimed to build a generic semantic middleware platform for smart cities. This unified the work conducted across separate projects in the second stage, and explored in more depth the 3rd and 4th research questions. A platform was developed which integrated several relevant components into a coherent API structure, including a SPARQL endpoint, a Hypercat endpoint, a KairosDB timeseries data endpoint, and BIM and CityGML endpoints. The platform reused libraries to provide the BIM and CityGML endpoints, but developed a library for storing and querying Hypercat item and catalogue objects. This library used a SetMultimap object to store Hypercat metadata about items, in Java, which is well suited to this purpose, as Hypercat metadata of both items and catalogues use an RDF-like approach to describing web resources. The internal structure of the BIM servlet was significantly simpler than the CityGML endpoint, primarily because the internal data structure used in the CityGML4J toolbox are significantly more complicated, in part because CityGML is centred on the concept of geometric features.

As well as the smart city server developed, a graphical interface was developed based on the CesiumJS engine. This application worked in tandem with the server to populate the interface with geometry and data about city objects as well as dynamic data from sensors. The static data about city objects, as well as links to their Things and services, are all stored in each entity's Cesium description. Whilst this does achieve the intended functionality, Cesium descriptions are HTML documents stored as plain strings, which was very cumbersome to embed callback functions into, as this aspect is primarily intended for plain or formatted descriptions of the objects, rather than dynamically created rich content and JavaScript. Future development could extend the Cesium library with 'Things' and 'Services' properties for entities, which are then handled independently by the infobox, outside of the description div element.

The use of semantic modelling techniques is a core benefit of the platform. This uses an ontology to formalise a description of the domain, which enriches the data with meaning and context. This goes beyond existing IoT platform approaches, in that the knowledge base includes rich semantic models of the built environment and its socio-technical systems as well as descriptions of the devices in the area. This allows the web services to reason over the meaning of the dynamic data in a much fuller sense, by applying AI to interoperable knowledge of the underlying systems in order to produce higher-order knowledge. The semantic knowledge base is deployed as a triple store, using the Apache Jena framework. The web services query the knowledge base with SPARQL to return static information directly, or to receive linked data URIs.

The platform integrates several features so as to provide a whole value-chain approach to empowering urban decision makers, as an extension of the current IoT, BIM, DST and KBS state of the art within a novel urban sustainability platform and interface. The solution uses an engaging 3D game engine to promote interactivity and responsiveness, and is closely coupled with the semantic web service and other back-end components, providing rich engagement with the virtual urban landscape. Finally, the system's interoperability approach emphasises the semantic description of the underlying socio-technical-cyber network as well as device and data knowledge; to expose a full and rich description of the environment to connected web services.

The proposed solution has the advantage over existing solutions of utilising its semantic web service, the ‘ontology server’, to describe the urban environment itself in a machine-interpretable manner, as opposed to current IoT platforms, which focus on describing the cyber aspects of smart connected devices. The platform is also intended to be highly extensible, such that new client applications can be developed as smart city use cases emerge, relevant research fields advance, and new human-machine technologies come to market. Having decoupled domain logic into a rich semantic model, these applications can be much thinner than if they had to cope with complex domain semantics. The front-end developed could be used as a template for incorporating additional functionality, or could be used alongside separate front-end components which provide intelligent analytic services.

## 5.2 OVERALL ACADEMIC CONTRIBUTIONS

This section now discusses the work conducted and outputs produced across the investigation, from the perspective of various fields of research. Firstly the contribution to the IoT field is discussed, before the contribution to smart city research in general, and finally to the main application domains targeted; smart energy and water systems.

### 5.2.1 IOT RESEARCH: INTEGRATING IOT AND DOMAIN SEMANTICS

The work conducted can be considered within the discourse of IoT research. The contribution of the work conducted is now discussed in relation to existing IoT platforms in research, the state of the art research around the semantic web of things, and briefly with regards to other pertinent research.

Firstly, it should be stated that most IoT platforms do not tackle semantic interoperability issues explicitly. Compared to those platforms which do not include semantic context for their data and services, the developed systems offer the previously mentioned benefits of semantic technologies. Some platforms do acknowledge the benefits of addressing semantics directly, such as the ALMANAC platform [18]. However, most SWoT ontologies consider only ICT concepts such as device status, services, and accuracies [269], [275], [319], [325], [584]. The platform developed in stage 3 offers the critical advantage of integrating ICT-domain descriptions of resources with rich application domain models for more powerful search and contextualisation. Also, none of the observed works used a standard API or response structure, whereas using the standard Hypercat format and API structure for resource discovery is beneficial. Also, offering various coherent programming and graphical interfaces for discovering resources and querying data supports the multidisciplinary nature of smart cities. Most previous research has followed an approach similar to ALMANAC, which refers to semantic data as 'metadata' [18]. However, the current work views semantic data as data in its own right, rather than just being used to supplement the data from the timeseries database. For example, water network GIS data was stored in a semantic web format in stage 2, and was retrieved by applications for direct visualisation, rather than being used as metadata. Emphasising the role of semantic data improves the value derived from its use, by promoting homogeneity amongst datasets and supporting the development of 'thin' applications.

Secondly, the work can be viewed within the discourse of the ongoing evolution of IoT towards a semantic web of things, as pioneered by the W3C Web of Things working group [277], and the Hypercat consortium [47]. Research towards a semantic web of things is embryonic, and regarding its semantic aspects, current work is exploring the data formats and ontologies necessary. The work conducted extends that of the Hypercat consortium by integrating the Hypercat API and response format with semantic web standards such as SPARQL and RDF. The integration of rich domain semantics with IoT semantics is also valuable, as it greatly facilitates the integration of data and services from different companies and domains, by reducing the ambiguity of the resource within the application domain. Specifically, this allows software developers to produce thinner applications, with greater confidence in data semantics, and also allows a greater depth of machine comprehension of the data and services, supporting semantic inference and other AI technologies in providing domain value.

Semantics and IoT are also being researched with regards to inference over data streams, which is particularly relevant with growth of big data. The work conducted didn't directly address stream data processing, although the ontologies produced could be used for this purpose. The approach adopted within the water domain investigation was to store the latest readings from each sensor in the triple store, and to apply inference over these values. This approach was sufficient for the intended purpose, but is not scalable to volumes and velocities of data which are orders of magnitude larger, and it precludes the inference of patterns in data streams, as it only considers the latest values. The ongoing work in the RDF stream data processing community would therefore be valuable and complementary to the conducted work.

### 5.2.2 SMART CITY RESEARCH: EMPOWERING DISCOVERABILITY AND BUILDABILITY

Within the discourse of smart city research, the work conducted furthers the state of the art from several perspectives. Firstly, the primary contribution to this field is empowering the ‘data hubs’ currently in widespread use [158], [585]. These data hubs are oriented around datasets, rather than the objects and data which they represent, which make it harder to discover and comprehend data through them. The platform developed overcomes this by focusing on the objects in the domain, and using those objects and their descriptions to contextualise the data from them. As well as facilitating the development of commercial software, this promotes the use of open data at the grassroots level by lowering the barriers of discoverability and comprehension of the data and services. By explicitly stating in OWL the context of datasets and streams, thin applications can be built on top of them with less technical knowledge and very little domain expertise. In an industrial IoT setting, this results in reduced development times, costs, and complexities (for extension and maintenance).

Another benefit of the developed platform is its multi-level API. This is a significant benefit as it allows data to be served in formats familiar to developers of different disciplines, rather than forcing a new toolset onto them. Specifically, developers familiar with BIM concepts or CityGML concepts can use the APIs of those standards to retrieve relevant data. Whilst this is not currently integrated with the triple store, that could be accomplished once semantic web versions of those standards are themselves standardised. Also, the upper ontology allows the platform to integrate data and services across smart city domains, which was not observed in the literature. The use of the Hypercat API and file format also aids in discoverability of resources, as it provides a standard and simple means to retrieve human and machine readable descriptions of the available resources. Finally, the

3D interface of the platform promotes engagement with the platform, and provides a proof of concept of how to leverage the platform, which could then be exploited by developers, and grassroots enthusiasts, to explore the available resources in a visual manner.

### 5.2.3 ENERGY AND WATER RESEARCH: SMART GRIDS NEED SMART DATA

This section discusses the work conducted in the 2nd stage, within the discourse of the application domains of the energy and water sectors. It is pertinent to discuss the work in these sectors alongside each other, as they share several similarities, and much can be learned from the energy sector and applied to the water sector both in research and practice.

Firstly, in the energy sector, at the building level, the solution space established within the literature has typically only explored the use of optimisation, simulation, or semantic web data integration independently, and mostly *in vitro*, as discussed in the literature review. The work conducted takes a step beyond the state of the art by addressing semantic interoperability in an explicit and automated way, and deploying this *in vivo*, demonstrating suitability of the technologies for retrofit in real world systems. The developed solution also combines the advantages of optimisation, simulation, rule-based systems, and data integration. For example, using an ANN model as a surrogate of the simulation model reduces the prediction time of the energy consumption and PMV variables. Finally, the combined use of empirical and theoretical rules allows a wide range of sensitive variables to be considered.

Next, at the smart energy grid level, the work conducted provided evidence in the domain of smart prosumer grids. The work was especially valuable as it served to explore the automation of several energy management decisions without requiring human authorisation or negotiation. This highlighted the breadth of ontological engineering required when applied to different domains, target systems, and for different usage patterns, despite using the same fundamental modelling language and constructs. The semantic artefacts performed sufficiently well for the target system's purposes, but their potential for reuse was reduced, due to the very close coupling of the ontology with the target ICT solution. This emphasised the importance of upfront decisions about the purpose of an ontology to be developed: the developers should agree and understand the implications of whether it is intended purely for one purpose or the degree of reuse which is desired from it. The use of semantic clarity was especially useful as the ICT solution integrated building-scale prosumer energy management with grid scale power management; M2M integration between applications and data in smart homes with grid-scale management schemes is highly valuable, and greatly assisted by semantic clarity. The same is true in integrating data and applications within homes and across grids of different utility-centric systems. An example of this would be the management of electrical and water appliances, including their nexus during heating water, or using appliances such as washing machines. Domestic ‘smart grid’ and ‘smart water’ applications at the property would hence have to interoperate if they were actuating or making suggestions about decision variables which affected the other system. Applications which successfully bridge the semantic complexity of integrating two expert technological domains to any non-trivial level of complexity would be arduous and potentially prohibitive without abstracting domain semantics from the application logic through semantic technologies.

Following from the lessons learnt in the energy sector, a significant undertaking was pursued of applying these to the water sector, which has little or none of the appreciation for the importance of semantics observed in the energy sector. In the water sector, it is common for system integration to be ad-hoc, and require a manual mapping between each heterogeneous component. As water networks become smarter, the time-intensive nature of this process, and of expert interpretation of the network’s data, will prohibit business-as-usual approaches. In order to overcome these challenges, alongside the trend of IoT, semantic modelling of the water industry must be undertaken. This need has been widely acknowledged in smart power grids through IEC standards [469] and in the building information modelling field through BuildingSmart [586], eeBUS [587], Haystack [588] and BSI standards. Of the various interoperability challenges, semantics have been particularly noted in the water sector, with the ICT4Water cluster of EU research projects noting “that semantics is the most important hurdle to overcome, even preceding the other priority sectors” [111].

The work conducted in the water sector successfully leveraged semantic technologies to improve the knowledge management of a water utility. Semantic models store knowledge about the network in a more comprehensive manner than traditional GIS or database-centric approaches, such that intended meaning is more precisely and reliably shared between interoperating software. Further, as relationships between data concepts are expressed, this enables the application of a knowledge based systems approach to the data management, whereby the water sector's decision makers are empowered by comprehensive, timely, and accurate knowledge, which is coupled closely with their management processes.

A knowledge-based approach to water system data offers several benefits. For example, rather than a water operations manager being notified of a pump alarm and then having to cross-reference and apply expert knowledge using several ICT systems, this can be achieved automatically, with semantics providing the mappings between systems to inform the expert; (a) exactly what the error is, (b) what the likely cause is, (c) what the impact is likely to be, and (d) what actions should be taken. The semantic web approach also presents the key benefit to utility companies of avoiding vendor lock-in, and is built from the ground-up to allow extensibility and scalability as the company's sensing and ICT infrastructure grows and changes over time. This technology also paves the way for enabling the use of artificial intelligence for the processing of water network data network, providing further added value to the data being collected. This inherent extensibility could also allow for an evolution whereby upgraded infrastructure and systems can infuse semantic meaning at lower levels of the technological stack thus moving towards increasingly distributed models of IoT. The semantic models of the WISDOM project represent a significant step towards this interoperable and knowledge-based approach.

A key contribution of the work conducted was the water domain ontology, as semantic modelling has been identified as a critical obstacle [111]. Whilst other ontologies take fundamental steps towards overcoming this obstacle, they are either not intended for the smart water domain, or are only suitable for capturing generic knowledge about a water network and relating observations to features of interest and alerts. The key novelty presented lies in the semantic representation of the water value chain as a detailed manifestation of a socio-technical-sensory system, at the network and building scales. This goes beyond the ontological modelling conducted elsewhere to offer greater depth and breadth. Specifically, the 'observation and measurement ontology' of the WatERP ontology is similar to the

WISDOM sensor ontology, due to their shared roots in the W3C SSN ontology [330], although the WatERP ontology's alignment with the SSN ontology is shallow; only reusing a few high level concepts. However, the WISDOM sensor ontology thoroughly reuses the SSN ontology, and extends it directly in order to be relevant to the water domain. The WatERP 'supply and demand ontology' contains concepts from across the rest of the WISDOM ontology, but again only captures high level concepts such as physical element types (storage, transfer, etc.) and a few types of actors (bulk water suppliers, consumers, regulators and water utilities). Hence, the WISDOM ontology is suited to a different purpose to that which the WatERP ontology achieves. Further, the WISDOM ontology captures domestic knowledge, so as to allow the integration of consumers within the water value data chain and hence contextualize smart meter and behavioural readings.

## 5.3 RELEVANCE TO PRACTICE

The final section of this chapter discusses the relevance of the work conducted, artefacts produced, and knowledge contributions to industry and practitioners. Firstly, the relevance to the field of informatics and enterprise systems is discussed, before a discussion pertinent to IoT practitioners, and finally smart cities and systems in a broader sense.

### 5.3.1 INFORMATICS AND ENTERPRISE SYSTEMS

Within informatics and enterprise systems, the work conducted has highlighted that i) there are many different types of semantic model, and just as many different arrangements for how to leverage these through semantic technologies. Also, ii) the up-front effort required to ontology-enable a system is sometimes prohibitive, but can be significantly mitigated. Finally, iii) requirements engineering and a clear prioritisation of expert engagement and 'ownership' is crucial in successfully building and leveraging ontologies. This section discusses these points in turn.

#### 5.3.1.1 CONTEXT DEPENDANT MANIFESTATIONS OF ONTOLOGIES FOR VARYING PURPOSES

Many different 'recipes' for using semantic technologies are observed in the literature, and have been explored through this thesis, both in terms of various interpretations of 'semantic model' as well as leveraging these artefacts in different ways. For example, using an ontology as a collection of metadata ‘tags’ to add to datasets is distinctly different to the use of an ontology to classify objects into various groups with specific meanings. Even within the use of ontologies as knowledge bases of IoT resources and their contexts within application systems, developers can choose to either store timeseries data within the triple store itself, or just link to the data in an external database.

The variety of possible mechanisms for leveraging semantic technologies highlights the importance of understanding clearly the purpose of incorporating semantics explicitly into a system. The decisions surrounding the ontology and the associated software should be led by business cases around the purpose and scope of the resulting software. This variety of ‘what semantic web looks like’ is also important in engaging stakeholders in the ontological processes, as preconceived notions of the purpose of a ‘semantic component’ of a system may negatively influence software development from a lack of understanding. Figure 12 illustrated a perspective on the various levels of semantic complexity possible in an ontology.

#### 5.3.1.2 MITIGATING THE UP-FRONT DEVELOPMENT BARRIER OF ONTOLOGIES

One criticism typically levied against ontologies is the upfront development barrier, in that developing an ontology can be prohibitively time consuming. In many cases this is a fair criticism, especially where the system doesn’t involve many intersections of domain perspectives, or the data is very tabular. However, in many cases, it is likely that the benefits of addressing semantics explicitly are worth the additional investment, especially if certain measures are taken to mitigate the upfront effort. This is especially true when future costs of extending the software are included. The first mechanism for mitigating the upfront effort is the reuse of existing ontologies, and especially middle and/or upper ontologies, as this can accomplish much of the work or provide a metamodel as a shortcut to the desired artefact. Adapting ontologies with similar structures or from similar domains can also reduce the effort required. To these ends, the smart city ‘upper’ ontology presented in this thesis can greatly reduce the effort required to develop an ontology for smart city applications. The role of accepted methodologies in ontology engineering such as the NeOn methodology [313] is also crucial, as these provide template processes to go through and offer structure and an evidence-based approach. Also, the role of standards work around ontologies is critical, as efforts to develop ontologies, build consensus, facilitate adoption, and finally standardise ontologies in a domain can then provide a benchmark 'off the shelf' solution to greatly shortcut the process. Finally, it is important to generally prefer lightweight ontologies during scoping and requirement setting stages, as a smaller scope will clearly require less work to fill, yet is often ignored in pursuing detailed and yet unnecessary scopes.

#### 5.3.1.3 THE IMPORTANCE OF INDUSTRIAL ENGAGEMENT, SOFT ENGINEERING, AND ROBUST REQUIREMENTS

Discussion has already been made around the importance of requirements engineering in general, but it is also critical in ontology engineering to balance the often conflicting drivers. Broadly speaking, the owners of an ontology's target system won't understand ontologies, and any ontology developed is only useful if it's adopted and applied, so people have to understand it and engage with its development early on. Specifically, balancing software requirements with industrial engagement is important: computer experts must be able to develop applications from the model easily, but domain experts must clearly understand the terminology and cope with the abstraction and reuse of accepted ontological modelling practices. The work presented in this thesis benefits in this aspect from formal, regular, multi-stakeholder engagements in the process, and an iterative process. By prioritising collaboration, and early domain expert engagement in the requirement setting, industrial experts gain more of a sense of ownership of the artefact.

As ontologies are an emerging field, the soft aspects of the process are also due significant consideration; requirement engineering for ontologies in fields where they are novel requires a careful balance of semantic accuracy and domain relevance. If the approach prioritises the ontological aspects of the task at the exclusion of the others, it may result in an artefact which is clear to semantic experts and which is logically optimal, but which appears far removed from the actual language used by domain experts. This results in the model only being relevant within the originally intended setting, with less likelihood for reuse and adoption. A balance should always be sought between software, knowledge, and reuse requirements, as well as pursuing expert engagement.

### 5.3.2 EMPOWERING IOT

IoT technologies are increasingly being recognised as transformative, but significant interoperability challenges remain and are under-acknowledged. This section first discusses the role of semantic technologies in achieving rich resource descriptions, discovery, and powering AI applications, before then discussing the suitability of semantic technologies within the trend of big data in enterprise systems.

#### 5.3.2.1 NEXT GENERATION IOT THROUGH MACHINE-INTERPRETATION OF DATA MEANING

Currently, IoT interoperability technologies are focused on enabling systems to communicate with each other, but little attention is being paid to ensuring that they are communicating effectively or accurately. The work presented in this thesis has shown that supplementing IoT systems with semantic context enables them to interoperate more effectively, towards a true cyber-physical system. Specifically, by decoupling the domain logic from application logic through semantic abstraction, less work is required to build powerful applications on top of the data and devices in a system. Ideally, the semantic context achieved through the ontology developed should integrate the cyber and target-domain semantics, by aligning an ontology of devices and cyber concepts with one describing the functions of these within the application domain, such as in the smart city ontology developed. Another benefit of such an approach is enhanced discoverability of IoT devices and services based on their real world relevance, and a significantly richer response describing discovered resources in an appropriate format, such as the Hypercat standard incorporated in the work conducted.

Whilst IoT is currently answering the question of getting data to applications and the cloud, this needs to evolve into application-layer interoperability. The applications need to speak the same languages and share domain perspectives, or at least map to the same intermediary in order to interact meaningfully. A great deal of the value of human communication is derived from the context of messages and message content, rather than the raw data exchanged, and as computers become more intelligent and human-like, it is logical for them to mirror this communication paradigm.

As machines progress towards AI and other advanced applications, they require more context around their input data, in order to exhibit more generic intelligence over that data. Little work has been done to date at the intersection between IoT and AI, but this represents a research field of unprecedented value. If machines were able to truly understand the capabilities and the data from an IoT system, and subsequently apply cutting edge intelligence techniques to this knowledge, they could take vast steps in improving the way industrial systems, and beyond, are controlled and monitored. In order to achieve this however, it is a prerequisite for the semantics of the data and systems to be available explicitly for the AI and advanced applications to use. The work conducted has demonstrated that semantic web technologies and OWL ontologies are suitable for this purpose.

#### 5.3.2.2 ONTOLOGIES AND KNOWLEDGE-BASED SYSTEMS IN THE ERA OF BIG DATA

As ontologies and knowledge-based systems typically use novel data storage software such as triple stores, it is pertinent to weigh the merits of these against conventional relational databases in terms of scalability, and other considerations for use in enterprise systems. Regarding scalability, Oracle have successfully tested a knowledge base with over a trillion triples [589], as have AllegroGraph [590], and many free options exist which can support billions of triples, which is likely to be enough for the majority of systems, if good practice is followed in designing the ontology in a lightweight manner.

Triple stores have many benefits over relational databases, the most important of which are i) standardised query language, ii) schema flexibility at runtime, iii) integration of data across graphs, and iv) inferencing can produce new knowledge beyond those inputted directly. Having a standardised and more powerful query language, SPARQL, makes it much easier to traverse data schemas, and means that any chosen triple store can be replaced if it is no longer suitable, without changing the way it is queried. Schema flexibility is another significant benefit of graph databases, as the contrary has been a drawback of relational databases for some time. Finally, it should be noted that semantic technologies add scalability to a system with data that is widely varied, as it allows the context of data to be understood automatically, rather than requiring human intervention to understand each variable.

### 5.3.3 MODERN SMART CITY AND SYSTEM MODELLING

This subsection discusses the role of modern file formats and semantic web concepts in built environment information modelling, before then discussing the need for semantic technologies to empower smart city data hubs across disciplines. Finally, this subsection discusses the need for standardisation work within the studied field.

#### 5.3.3.1 BENEFITS OF MODERN KNOWLEDGE MODELLING LANGUAGES IN DIGITISING AECFM

The two main emerging data formats in smart cities are IFC and CityGML, although other formats are pertinent such as gbXML. Unfortunately these formats, and most other standards, are based on legacy languages, such as IFC being based on STEP-EXPRESS, and CityGML being an extension of XML. Simpler conceptual modelling technologies such as EXPRESS are less expressive than the W3C RDF model, and not using IRIs to refer to named individuals makes them less well-suited to use in web-enabled systems.

The work conducted in the building-energy domain utilised the emerging ifcOWL semantic web format, and showed that it is ideal for integration of web services. The domain ontology reused ifcOWL, and extended it with relevant sensing, actuation, energy, and behavioural concepts. Also, the work conducted in the water and energy grid domains aligned their ontologies with the IFC model, for future integration with ifcOWL when standardised. The benefit of the ifcOWL alignment, as well as SAREF alignment, is that these common languages facilitate interoperability with other existing or future smart home applications, allowing the data to be reused in a contextualized and meaningful manner. For example, water data from a washing machine which is used by the WISDOM app could be integrated with energy and scheduling data which is used by the MAS2TERING app, allowing optimization across both systems, due to the use of a common vocabulary. The work conducted in the 3$^{rd}$ stage incorporated CityGML and IFC APIs, although it didn’t integrate the underlying conceptual models with the smart city ontology because suitable standardised semantic web versions were not available.

The use of semantic web technologies to provide semantic context to data instead of legacy formats greatly facilitates the integration of static system descriptions with dynamic timeseries data from sensors, as the timeseries data can either be also stored in the triple store, or richly linked contextualised whilst being stored in an external database, as demonstrated through the work conducted in the WISDOM project and the smart city platform developed.

#### 5.3.3.2 UNLOCKING THE VALUE OF MULTIDISCIPLINARY LINKED OPEN DATA

One common feature of most smart city initiatives is an open data portal or 'hub', where local public bodies make a number of datasets available. These portals are typically focused on retrieving datasets, with limited tagging of the datasets in terms of their contents, and possibly regarding their quality. This limits the utility of the uploaded data, as people are more likely to perceive urban environments in terms of the objects within them, and are unlikely to be familiar with the semantics or terminology of the uploaded datasets. By making a common semantic framework for open data about smart cities, and allowing data and their semantics to be searched and contextualised, these hubs are likely to be far more useful to potential commercial projects, private usage, and grassroots app development. Also, by relating data from across organisations and domains to a coherent semantic framework, relationships across these systems can be found and more complex analyses are feasible, promoting a more intelligent system of systems approach to smart city management.

#### 5.3.3.3 THE ROLE OF STANDARDISATION IN URBAN SEMANTICS

Throughout the work conducted the importance of information modelling standards work has been highlighted. Ontologies were developed for the smart water and energy domain, as well as an upper level smart city ontology. Throughout these, the ability to reuse existing models and to use standards as a benchmark of expert consensus has been invaluable in reducing the work required, building confidence in the ongoing process and in the produced artefacts prior to formal validation. However, the need for further work, especially with a focus on rich semantics models for use in web contexts, cannot be understated. This work should prioritise rapid development of models and building of consensus within a clear scope and towards clear use cases, as it is not necessary or even ideal for one ontology to be universally adopted to benefit from consensus amongst a community.

The need for standards work was especially noted in the water sector, where significantly less pertinent standards development organisations are present than the energy sector. A lot of ground has been covered by organisations like the European Commission, NASA and the Open Geospatial Consortium, and by projects such as WatERP, SWIM and WISDOM. These have built a strong foundation for syntactic and semantic interoperability in the water sector, but there is a lot left to do. Critically, these largely separate efforts should be brought together in a coherent manner by the water industry. This will deliver a semantic model which is necessary for the next generation of truly smart water networks. The model will enable powerful interoperability across traditional silos in the water industry. The initiative has to be driven by industry, through a mechanism which echoes the transformation the construction industry has recently undergone with BIM through buildingSMART.

This chapter has provided a discussion around the evidence produced throughout the investigation, with the aim of exploring its knowledge contribution though the lens of the stated research questions and hypothesis. This has considered each action and design research iteration in depth, before unifying these threads of work and comparing the work and findings to the literature, and finally discussing the relevance to practice. The following conclusion chapter draws together this discussion to directly address the stated hypothesis and research questions. This succinctly outlines the main research findings, the novel contribution to the body of knowledge, and the limitations and potential further work.

# 6 CONCLUSION

Unifying the threads of discussion provided in the previous chapter, this chapter provides a summary of the work conducted and its main findings, framed through the research questions posed at the beginning of the investigation. The chapter first considers each of the research questions in turn, before discussing the limitations and potential for future work, and finally offering concluding remarks. The overarching perspective adopted is a heirarchical view of the academic contributions across the stages of the investigation. Specifically, the most important and unified contribution is derived from analysing across the entire investigation, and the next level of contribution significance relates to the work conducted in the 3rd stage of the investigation. The work of the 2nd stage then supported the higher tiers of contribution through a rigorous and broad evidence gathering and iterative learning process. The furthest removed stage from the central findings was then the 1st stage, which provided a broad literature review and theoretical consideration of the problem space, but provided little direct academic contribution.

## 6.1 MAIN RESEARCH FINDINGS

The objective of the investigation conducted was to explore the stated hypothesis through its decomposition into 4 research questions. The hypothesis explored was:

*A Semantic Web of Things approach to technology interventions can deliver value to smart city stakeholders by better leveraging IoT and AI synergistically to provide better decision support.*

This section restates and offers a concluding discussion for each of the research questions in turn, and discusses the pertinent work conducted, evidence produced, and main findings related to each.

### 6.1.1 SMART CITY THEORY AND REQUIREMENTS

The first research question posed was:

*What are the theoretical underpinnings of ICT knowledge gaps in smart cities, including the challenges, impact scenarios and scope for step changes?*

The work conducted through the literature review in the 1st stage of the methodology, and through the requirements engineering and expert engagement tasks of the 2nd stage, were the most pertinent to this research question. The literature review formed a theoretical underpinning for the research based on system theories, and elicited knowledge gaps from state of the art research in smart city informatics, IoT software, and applied semantic technologies. The literature review also began an ongoing effort towards application domain relevance by reviewing the latest research in smart energy and water grids. Overall, this painted a clear picture of a gap in the literature which pointed to a need for further research on semantic interoperability at the intersection of IoT and AI technologies in enterprise systems for smart city applications. Specifically, applied ontology research is critical, including development, application, and consensus building around ontologies and their supporting software, to produce reusable ontologies and evaluate the best approaches to leveraging them in IoT-AI systems.

Extending the work conducted through the literature review, the 2nd stage of the investigation involved engaging with several projects alongside experts. By consulting closely with many stakeholders in the technology value chain of smart cities, a better understanding was elicited of the practical nature of the semantic interoperability gap. Detailed requirements engineering processes were undertaken at the building and grid level in both the water and energy domains. This produced many scenarios and use cases for the application of semantic technologies, as well as comprehensive requirement statements in order to realise these. This work was also supplemented by generalising the energy and water domain research into a broader smart city consideration, which paved the way for the smart city semantic platform developed in the 3rd stage of the investigation.

It is concluded that a system of systems perspective is critical in optimally operating built environments. Specifically, if intelligence and even control of industrial systems is to be awarded to virtual agents rather than human agents, the communication between these agents should be supported in at least as rich and expressive a manner as would be expected between humans. Within human communication, the context and semantics of exchanged messages are crucial to enabling rich interactions and truly facilitating emergent behaviour, and this is equally true between machines, so should be automatically supported as such. It is essential that work is conducted towards these aims with a pragmatic stance of delivering value to the organisations and individuals involved in the target domains, so as to avoid the perceived ‘academic ambiguity’ of semantic technologies observed in stakeholders throughout the investigation.

### 6.1.2 INTEGRATING IOT AND AI THROUGH SEMANTIC TECHNOLOGIES

The second research question posed was:

*How can a semantic web of things approach integrate IoT and advanced smart city applications?*

The 3rd stage developed a software platform which aimed to answer this question based on the learnings and artefacts of the 2nd stage, and this software is detailed in the results section. This platform was based on the evidence gathered across the research projects engaged with through the 2nd stage of the investigation. Each project provided different use cases, stakeholders, and legacy systems; so a breadth of pertinent evidence was produced. Within the KnoHolEM project, the knowledge base was required to store many business rules, and so a triple store was integrated with a Jess rule engine and intelligent components which produced the rules. In the MAS2TERING project, an OWL ontology was integrated into a MAS to facilitate agent-based communication through a metaprogramming conversion to a JavaBeans format, and was also exposed through a triple store to integrate with a prediction service. In the WISDOM project, integration with legacy systems and timeseries data was critical, so the ontology was designed to link sensors with their timeseries data in a KairosDB server. These requirements also led to the developed system incorporating a message bus, and advanced inference capabilities through the Pellet and Drools engines.

It is concluded that many different arrangements of semantic technologies can integrate IoT and advanced applications within the smart city domain, each of which has its own characteristics and use case suitability. Several of these arrangements have been explored and demonstrated through this investigation to produce a better understanding of their implementation for novel applications. Semantic Web of Things approaches to using semantic technologies can be conceived as consisting of 3 parts: i) a set of knowledge, which is typically comprised of a domain ontology and an instance of this, ii) software components which store and manage the knowledge internally, and iii) a programming interface which exposes some functionality of the internal software to external applications. By varying the nature and use of these parts, especially the domain ontology, the performance and system integration functions of the service can be adapted to provide the necessary support to the broader system.

It was found, for the systems explored, that semantically-enabling an IoT system should ideally incorporate a comprehensive knowledge base; describing the Things in terms of their cyber and application domain contexts, as well as pertinent non-web-enabled objects in the system. This allows software, including AI, to be built on top of the available data and services in a much ‘thinner’ manner, whereby the domain logic is disaggregated from the application logic. This allows software to be developed more easily, and with more confidence and less complexity in the use of the domain data, services, and devices.

### 6.1.3 VALUE PROPOSITION OF SEMANTIC TECHNOLOGIES

The third research question posed was:

*What value does a semantic web of things approach offer technology providers and decision makers in smart city systems?*

As with the 2nd research question, a breadth of data was gathered through the 2nd research stage, and used to solidify the value proposition through design science in the 3rd research stage. Whilst the 3rd stage directly demonstrated the potential of a SWoT platform, the 2nd stage involved a qualitative analysis of the various artefacts and systems produced, a quantitative analysis of their performance, and consultations with expert partners to determine the pragmatic value to their organisation.

The value proposition of semantic technologies to smart city IoT systems depends on the target system, but a typical value pathway is now provided. Firstly, semantic technologies provide a powerful discovery service for web resources. Provided a web-enabled Thing has a useful URL which can be meaningfully dereferenced, an RDF triple store is an ideal way to provide a list of available resources. Secondly, this ‘list’ can be enhanced with an almost unlimited amount of contextualisation, to support more powerful semantic search of resources in large systems, and to provide a comprehensive description of discovered resources for software developers. This point stands regardless of whether the Thing has a dereferencable URL. Thirdly, this context can (and should) include application domain semantics in a completely integrated and flexible manner, to provide further richness to the description, and to explicitly indicate the role of the Thing within the target system. Fourthly, this contextualisation can be enhanced by using an inference engine, and possibly encoding expert knowledge into SWRL rules to supplement the potential inference. Finally, rich machine interpretable picture of a system and its parts allows application developers to discover and build on top of the available data in less time, towards more complex applications, and with more confidence in the assumed meaning of data.

It must also be mentioned that from a purely computer science perspective, triple stores offer several advantages over traditional relational databases, regardless of whether a full semantically-enabled approach is adopted. The main benefits of triple stores are that they offer a standard, powerful query language, schema flexibility at runtime, and the ability to query over multiple graphs seamlessly.

### 6.1.4 GENERIC SMART CITY SEMANTIC PLATFORM

The fourth research question posed was:

*Can these learnings and artefacts be generalised to support further work across smart city domains and semantic web of things research?*

This final question was primarily addressed through work conducted in the 3rd stage of the investigation. This involved a design research process to build a semantic smart city framework, with the aim of unifying the previous learnings and artefacts, and to facilitate further ontological modelling. To this end, a generic 'smart city server' was built which incorporated a triple store along with a coherent API for querying other built environment semantic model formats and timeseries data in an integrated manner. Further, the ontological modelling of system theories conducted in stage 2 was unified and extended into an upper smart city ontology, which facilitates extensions into other domains through metamodeling. In this way, using semantic technologies in any smart city domain can be integrated by incorporating conceptual abstraction (as is best practice) into the modelling of the domain, and then aligning the resultant ontology with the presently developed smart city ontology. This was conducted for the energy and water domain ontologies, to demonstrate the approach.

The system of systems and sociotechnical systems nature of smart city systems can be captured through ontological modelling, and then reused in other domains which exhibit systemic characteristics. Also, within urban environments, several features are likely to be relevant regardless of the application domain, such as buildings, green spaces, water bodies, transport artefacts, devices, pipes and cables: these have been captured in the developed ontology for extension into application domains. As well, the best practices and learnings discussed with regards to the 3rd research question are relevant regardless of the application domain. However, it must be stressed that the manner in which semantic technologies are applied in a target system should ideally be specific to that system, rather than adopting a generic system architecture and modelling paradigm.

Whilst the developed smart city ontology successfully acts as a proof of concept, and can be used as a reference point for further ontological modelling, it does not currently represent a consensus amongst experts of how smart cities are perceived, or should be modelled. The role of standardisation work around semantic technologies cannot be understated, as the value of ontologies is closely linked to the degree of consensus and adoption they represent. This leads to the beginning of recommendations for further work.

### 6.1.5 REVISITING THE RESEARCH HYPOTHESIS

The initial hypothesis of the investigation was:

*A Semantic Web of Things approach to technology interventions can deliver value to smart city stakeholders by better leveraging IoT and AI synergistically to provide better decision support.*

This assertion was tested by investigating the 4 research questions posed. The 1st research question elicited a rigorous grounding and conceptual framework for understanding the hypothesis' assertion properly. Specifically, the perceived natures and roles of IoT, AI, and semantic technologies were clarified, and what delivering value would require in the smart city domain. From this, it was understood that semantic technologies should promote a system of systems approach to managing the built environment, primarily by empowering inter-system communication, in order for the investigation to support the hypothesis.

The 2nd research question demonstrated various possible value pathways of semantic technologies. This showed that semantic technologies can play a role in leading-edge innovations in smart systems. Specifically, this showed the various ways semantic technologies can promote synergy between the IoT and AI aspects of ICT interventions in industrial systems through a knowledge-based approach including application-domain knowledge.

The 3rd research question then explored in more depth the value brought to the smart systems engaged with by using a semantic approach, which forms the knowledge most pertinent to the hypothesis. The value proposition of semantic technologies to smart city stakeholders are manifold, as discussed previously. To summarise, adopting a semantic approach promotes powerful IoT resource discovery and contextualisation, allows integration with inference engines and allows applications to be much thinner.

The 4th research question then generalised the prior outputs to a broader range of ‘smart city stakeholders’. Through metamodeling, and the development of generic software platforms, the value of semantic technologies is increased as components can be reused both directly and indirectly to reduce future barriers to adopting a semantic approach. It also became clear that the manner in which semantic technologies should be leveraged in a target system can vary significantly depending on the system, its application domain, and the stakeholders’ objectives. It also became clear that semantic technologies are not a ‘silver bullet’, and are not suitable in all situations.

Overall, the main value of semantic technologies lies in mitigating the semantic interoperability barriers which arise when two or more agents communicate data to each other. Currently, these barriers are overcome manually in an ad-hoc manner, through discussion and expert investigation between ICT and domain professionals and the target system. However, as the scale of IoT penetration grows, heterogeneous systems will increasingly be expected to interoperate, as ICT systems in general become more and more interconnected. The scale of the challenge of achieving semantic interoperability in a system increases with the number of agent interactions, the heterogeneity of the agents’ domain perspectives, the complexity of the domain or datasets, the criticality of the target system, and the quantity of exchanged knowledge at each interaction. This is pertinent because the semantic challenge in systems is generally increasing as IoT interconnects more and more broadly, and semantic technologies greatly reduce these challenges.

Specifically, semantic technologies offer value where semantic interoperability may otherwise have significantly negatively impacted the investment required, time to market, functionality, or confidence in an ICT system, or is likely to in the future. Pragmatically, if a system is simple or clearly bounded to a few agents with similar domain-perspectives, incorporating semantic technologies is unlikely to make business sense. However, many smart city undertakings are now involving multiple stakeholders with varying perspectives, and expect software to interoperate in a system-of-systems nature, where semantic technologies can bring much value.

## 6.2 KEY CONTRIBUTIONS TO THE BODY OF KNOWLEDGE

This section describes the core contribution of the investigation and its relevance in furthering the body of knowledge. This provides an overarching perspective of the main research findings, based on a hierarchical view of the contributions from the learning iterations undertaken. From this view, the top level of the contribution hierarchy is the overall finding of the value of a Semantic Web of Things for smart cities. This is supported strongly by the next level of the contribution: the novel and significant work conducted in developing a smart city server, ontology, and GUI in the 3$^{rd}$ research stage (presented in Section 4.4). The rigour of the 3$^{rd}$ research stage was then supported by the extensive work conducted through the 2$^{nd}$ research stage, consisting of multiple learning iterations of action research (presented in Section 4.2 and Section 4.3). The lowest level of the contribution heirarchy is then finally the 1$^{st}$ research stage, which provided a comprehensive understanding of the problem space and its conceptual grounding (presented in Section 4.1 and supported by the literature review in Section 2).

The key contribution of the investigation is knowledge of the value and means of progressing towards a Semantic Web of Things for smart cities. Specifically, the work demonstrated that enriching IoT devices, services, and data with semantic web descriptions effectively evolves the Internet of Things to a Semantic Web of Things, which better supports applications and intelligent agents in managing the built environment. Specifically, the use of a SWoT approach allowed rich knowledge of the context of devices to be shared amongst business services, which greatly promoted interoperability. This scalable interoperability promoted an integrated consideration across previously siloed systems, and moved towards leveraging 'system of systems theories', to optimise built environment management more holistically. This was achieved by improving application-layer IoT communication. Within the case studies engaged with, this i) allowed facility managers to reduce energy consumption by circa 30%, ii) enabled water utilities to better manage regulatory compliance regarding CSO spill events, and iii) enabled applications to effectively consider the water-energy nexus by overcoming heterogeneity between the two domains, amongst other positive results.

The investigation explored the problem space iteratively towards a final learning cycle, which built on the rigour of the broad investigation and developed a novel software platform, ontology, and GUI, in the 3$^{rd}$ stage of the methodology. The software platform proposed an integrated suite of APIs for accessing semantically-enriched built environment data from various perspectives. This leveraged a prevailing IoT interoperability standard in a unique way, by building a Java binding for the data format and its implicit model, and using this to extract and transform device metadata from a triple store into the appropriate JSON serialisation. The API also exposed a full SPARQL endpoint for rich querying of the data, as well as BIM, CityGML, and timeseries endpoints for specifically accessing those views of the data. Coupled with the comprehensive ontological modelling conducted of the smart city domain, this provides a highly extensible means of providing rich cyber-physical semantics to IoT data, and hence replicating the positive results observed from leveraging a SWoT approach through this investigation.

As well as supporting the outputs of the 3$^{rd}$ research stage, the 2$^{nd}$ research stage contributed a rigorous breadth of evidence itself. This was achieved by exploring the problem space thoroughly within the building, energy and water domains, within a wide range of systems, use cases, and software architectures, and alongside a range of expert stakeholders. This variety of experimentation provided a wealth of qualitative and quantitative data by which to triangulate findings and evaluate the research hypothesis, where each iteration provided preliminary learnings and supported the later stages of work. The lowest level in this hierarchy of contribution was then the 1$^{st}$ research stage, which thoroughly explored the literature surrounding the various fields which SWoT research intersects with, and used an analysis of existing theories and approaches to postulate a conceptual model for SWoT systems.

## 6.3 LIMITATIONS AND FUTURE WORK

The investigation was conducted over the course of 3 years, involving engagement with 6 research projects and circa 40 organisations across stakeholder perspectives. However, the research and hence findings are subject to a number of limitations. One primary limitation of the work is that the research projects engaged with were not fully subject to typical business drivers as they were supported by research grants. Whilst the cross-sectional qualitative consideration of the developed artefacts and systems is not weakened by this, greater depth could be explored regarding the longitudinal impact of semantic technologies in enterprise systems. This should explore in more depth the value which semantic technologies offer organisations with regards to their key performance indicators and business drivers.

Another clear limitation of the investigation is the rapidly evolving nature of the studied field. Specifically, IoT research, adoption, and standardisation, has undergone transformative change over the course of the investigation. Due to this, new standards and technologies have become available and expert perceptions have changed between the stages of the investigation, which is not conducive for academic rigour, albeit unavoidable and not a critical weakness. This transformative change is likely to continue for the near future, but further work could continue to explore the role of semantic technologies with this knowledge and incorporate it into the methodological design.

The final limitation to mention is the breadth of the work across smart city domains. Specifically, the work only conducted action research across projects in the energy and water domains, before then generalising to the smart city upper domain in the 3rd stage of the investigation. The impact of this limitation was mitigated by reusing and aligning with ontologies adopted across the range of smart city domains, and consulting with organisations with expertise across smart city domains, such as local municipalities. Despite this, the suitability of the ontological modelling for broader domains has not been tested. The overall findings are unlikely to be affected by this, as the IoT stack has been applied broadly across all smart city domains, but further work could explore the applicability of the findings to other domains such as mobility, health and security/crime.

As mentioned, the role of standards in leveraging semantic technologies is critical, as their value is partly derived from their representation of a domain consensus. Therefore, further work could be conducted towards building such a domain consensus around the ontologies produced in this investigation. Especially, further research should investigate expert domain perceptions around the scope of the smart water and smart city ontologies developed, as these could fill significant gaps in both academic literature and in practice. Detailed, comprehensive and systematic consultation processes could be undertaken to elicit constructive feedback in iterative cycles until a true consensus is reached and the models can be relied upon as de facto standards.

## 6.4 CLOSING REMARKS

This thesis has applied semantic technologies at the intersection of IoT and AI technologies within smart city systems. Smart energy, water and city ontologies have been developed, as well as accompanying software, and the value of incorporating semantic technologies into urban ICT interventions has been demonstrated and explored. A great deal of work remains before a true semantic web of things may be achieved, but the work conducted acts as a proof of concept of the approach and provides a number of artefacts and knowledge contributions which can assist the ongoing research and development effort.

# 7 REFERENCES

[1] United Nations, "Smart Cities: Regional Perspectives." 2015.

[2] BSI, "Making cities smarter: Guide for city leaders," 2015.

[3] M. Dodgson and D. Gann, "Technological Innovation and Complex Systems in Cities," *J. Urban Technol.*, vol. 18, no. 3, pp. 101–113, Jul. 2011.

[4] M. L. Tuballa and M. L. Abundo, "A review of the development of Smart Grid technologies," *Renew. Sustain. Energy Rev.*, vol. 59, pp. 710–725, Jun. 2016.

[5] A. Cocchia, "Smart and Digital City: A Systematic Literature Review," in *Smart City*, R. P. Dameri and C. Rosenthal-Sabroux, Eds. Cham: Springer International Publishing, 2014, pp. 13–43.

[6] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in Europe," *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, 2011.

[7] BSI, "Smart Cities overview – Guide," BSI:PD 8100, 2015.

[8] C. Harrison *et al.*, "Foundations for Smarter Cities," *IBM J. Res. Dev.*, vol. 54, no. 4, pp. 1–16, Jul. 2010.

[9] P. Brizzi, D. Bonino, A. Musetti, A. Krylovskiy, E. Patti, and M. Axling, "Towards an ontology driven approach for systems interoperability and energy management in the smart city," in *2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, 2016, pp. 1–7.

[10] R. A. Alshawish, S. A. Alfagih, and M. S. Musbah, "Big data applications in smart cities," in *Engineering & MIS (ICEMIS), International Conference on*, 2016, pp. 1–7.

[11] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," *J. Internet Serv. Appl.*, vol. 6, no. 1, Aug. 2015.

[12] G. Gardašević *et al.*, "The IoT Architectural Framework, Design Issues and Application Domains," *Wirel. Pers. Commun.*, Oct. 2016.

[13] A. Sajid, H. Abbas, and K. Saleem, "Cloud-Assisted IoT-Based SCADA Systems Security: A Review of the State of the Art and Future Challenges," *IEEE Access*, vol. 4, pp. 1375–1384, 2016.

[14] AIOTI, "Semantic Interoperability Release 2.0," Internet Architecture Board, 2015.

[15] IEEE Standards Committee, P. IEEE Standards Coordinating Committee 21 on Fuel Cells Dispersed Generation and Energy Storage, Institute of Electrical and Electronics Engineers, and IEEE-SA Standards Board, *IEEE guide for smart grid interoperability of energy technology and information technology operation with the electric power system (EPS), end-use applications and loads*. New York, N.Y.: Institute of Electrical and Electronics Engineers, 2011.

[16] BSI, "Smart city concept model- guide to establishing a model for data interoperability." 2014.

[17] D. Remund and D. D. Aikat, "Drowning in Data: A Review of Information Overload within Organizations and the Viability of Strategic Communication Principles," in *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*, 2012, pp. 231–250.

[18] D. Bonino *et al.*, "ALMANAC: Internet of Things for Smart Cities," 2015, pp. 309–316.

[19] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An Information Framework for Creating a Smart City Through Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 112–121, Apr. 2014.

[20] James Manyika *et al.*, "Unlocking the potential of the internet of things." McKinsey Global Institute, 2015.

[21] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[22] UK Government Chief Scientific Advisor, "Internet of Things: making the most of the Second Digital Revolution." The Government Office for Science, 2014.

[23] A. Sheth, "Internet of Things to Smart IoT Through Semantic, Cognitive, and Perceptual Computing," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 108–112, 2016.

[24] W. Wang, S. De, G. Cassar, and K. Moessner, "Knowledge representation in the internet of things: semantic modelling and its applications," *Autom. Control Meas. Electron. Comput. Commun.*, vol. 54, no. 4, 2013.

[25] X. Su, J. Riekki, J. K. Nurminen, J. Nieminen, and M. Koskimies, "Adding semantics to internet of things: ADDING SEMANTICS TO INTERNET OF THINGS," *Concurr. Comput. Pract. Exp.*, vol. 27, no. 8, pp. 1844–1860, 2014.

[26] N. Zhang, H. Chen, X. Chen, and J. Chen, "Semantic Framework of Internet of Things for Smart Cities: Case Studies," *Sensors*, vol. 16, no. 9, p. 1501, Sep. 2016.

[27] D. Pfisterer *et al.*, "SPITFIRE: toward a semantic web of things.," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 40–48, 2011.

[28] A. Gyrard, C. Bonnet, and K. Boudaoud, "Domain knowledge Interoperability to build the Semantic Web of Things," 2014.

[29] A. Gyrard, M. Serrano, and G. A. Atemezing, "Semantic web methodologies, best practices and ontology engineering applied to Internet of Things," in *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, 2015, pp. 412–417.

[30] J. Vázquez Salceda *et al.*, "Making smart cities smarter using artificial intelligence techniques for smarter mobility," in *SMARTGREENS 2014:*

*proceedings of the 3rd International Conference on Smart Grids and Green IT Systems*, 2014, pp. IS7–IS11.

[31] E. Kuznetsova, Y.-F. Li, C. Ruiz, and E. Zio, “An integrated framework of agent-based modelling and robust optimization for microgrid energy management,” *Appl. Energy*, vol. 129, pp. 70–88, Sep. 2014.

[32] PS Satsangi, DS Mishra, Sant Gaur, BK Singh, and DK Jain, “Systems dynamics modelling, simulation and optimization of integrated urban systems: A soft computing approach,” *Kybernetes*, vol. 32, no. 5/6, pp. 808–817, 2003.

[33] A.-T. Nguyen, S. Reiter, and P. Rigo, “A review on simulation-based optimization methods applied to building performance analysis,” *Appl. Energy*, vol. 113, pp. 1043–1058, 2014.

[34] CityPulse, “citypulse_d5.2: User-Centric Decision Support in Dynamic Environments.” 2015.

[35] M.-A. Aufaure, R. Chiky, O. Curé, H. Khrouf, and G. Kepeklian, “From Business Intelligence to semantic data stream management,” *Future Gener. Comput. Syst.*, vol. 63, pp. 100–107, Oct. 2016.

[36] A. Chianese, P. Benedusi, F. Marulli, and F. Piccialli, “An Associative Engines Based Approach Supporting Collaborative Analytics in the Internet of Cultural Things,” 2015, pp. 533–538.

[37] H. Baars and J. Ereth, “FROM DATA WAREHOUSES TO ANALYTICAL ATOMS–THE INTERNET OF THINGS AS A CENTRIFUGAL FORCE IN BUSINESS INTELLIGENCE AND ANALYTICS,” 2016.

[38] M. Cavada, D. V. Hunt, and C. D. Rogers, “Smart cities: Contradicting definitions and unclear measures,” in *World Sustainability Forum*, 2014, pp. 1–12.

[39] V. Albino, U. Berardi, and R. M. Dangelico, “Smart Cities: Definitions, Dimensions, Performance, and Initiatives,” *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015.

[40] K. Su, J. Li, and H. Fu, “Smart city and the applications,” in *Electronics, Communications and Control (ICECC), 2011 International Conference on*, 2011, pp. 1028–1031.

[41] Anthony G.O. Yeh, “Making Intelligent and Smart Cities SMARTER,” presented at the CUPUM 2013, Utrecht, The Netherlands, 2013.

[42] M. Batty *et al.*, “Smart cities of the future,” *Eur. Phys. J. Spec. Top.*, vol. 214, no. 1, pp. 481–518, Dec. 2012.

[43] M. Zdravković, F. Luis-Ferreira, R. Jardim-Goncalves, and M. Trajanović, “On the formal definition of the systems’ interoperability capability: an anthropomorphic approach,” *Enterp. Inf. Syst.*, pp. 1–25, Jun. 2015.

[44] Shi-Wan Lin and Bradford Miller, “Industrial Internet: Towards Interoperability and Composability,” Industrial Internet Consortium, Jun. 2016.

[45] A. Gyrard, M. Serrano, and P. Patel, “Building Interoperable and Cross-Domain Semantic Web of Things Applications,” *ArXiv Prepr. ArXiv170301426*, 2017.

[46] S. Mayer, J. Hodges, D. Yu, M. Kritzler, and F. Michahelles, “An Open Semantic Framework for the Industrial Internet of Things,” *IEEE Intell. Syst.*, vol. 32, no. 1, pp. 96–101, 2017.

[47] Hypercat Limited, “BSI:PAS 212:2016 Automatic resource discovery for the Internet of Things.” BSI, 2016.

[48] European Research Cluster on the Internet Of Things, “IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps,” Mar. 2015.

[49] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993.

[50] J. Davies, R. Studer, and P. Warren, *Semantic Web technologies: trends and research in ontology-based systems*. Chichester, England; Hoboken, NJ: John Wiley & Sons, 2006.

[51] M. Uschold and M. Gruninger, “Ontologies: Principles, methods and applications,” *Knowl. Eng. Rev.*, vol. 11, no. 02, pp. 93–136, 1996.

[52] N. F. Noy and D. L. McGuinness, *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001.

[53] G. F. Luger, *Artificial intelligence: structures and strategies for complex problem solving*. Pearson education, 2009.

[54] J. McCarthy, M. L. Minsky, N. Rochester, and Claude E. Shannon, “A Proposal for the Dartmouth Summer Research project on Artificial intelligence.” Dartmouth College, 1955.

[55] A. Turing, “Computing Machinery and Intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[56] UCI-ICS, “Introduction to AI & Intelligent Agents,” University of California, 2016.

[57] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Englewood Cliffs, N.J: Prentice Hall, 1995.

[58] M. E. Bratman, D. J. Israel, and M. E. Pollack, “Plans and resource-bounded practical reasoning,” *Comput. Intell.*, vol. 4, no. 3, pp. 349–355, 1988.

[59] President’s Council of Advisors on and Science and Technology, “Technology and the Future of Cities,” 2016.

[60] M. E. El-Hawary, “The smart grid - State-of-the-art and future trends,” *Electr. Power Compon. Syst.*, vol. 42, no. 3–4, pp. 239–250, 2014.

[61] S. M. Amin and B. F. Wollenberg, “Toward a smart grid: power delivery for the 21st century,” *Power Energy Mag. IEEE*, vol. 3, no. 5, pp. 34–41, 2005.

[62] V. Giordano *et al.*, “Smart Grid projects in Europe,” *JRC Ref. Rep. Sy*, vol. 8, 2011.

[63] X. Fang, S. Misra, G. Xue, and D. Yang, “Smart Grid - The New and Improved Power Grid: A Survey,” *IEEE Commun. Surv. Tutor.*, vol. 14, no. 4, pp. 944–980, 2012.

[64] M. Ahat, S. ben amor, and A. Bui, “Agent Based Modeling of Ecodistricts with Smart Grid,” in *Advanced Computational Methods for Knowledge Engineering*, vol. 479, 2013, pp. 307–318.

[65] P. J. Werbos, “Computational Intelligence for the Smart Grid-History, Challenges, and Opportunities,” *IEEE Comput. Intell. Mag.*, vol. 6, no. 3, pp. 14–21, Aug. 2011.

[66] H. Jiayi, J. Chuanwen, and X. Rong, “A review on distributed energy resources and MicroGrid,” *Renew. Sustain. Energy Rev.*, vol. 12, no. 9, pp. 2472–2483, Dec. 2008.

[67] W-D. Zheng and J-D. Cai, “A multi-agent system for distributed energy resources control in microgrid,” 2010, pp. 1–5.

[68] Z. Jiang and R. A. Dougal, “Hierarchical microgrid paradigm for integration of distributed energy resources,” in *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008, pp. 1–8.

[69] M. Manfren, P. Caputo, and G. Costa, “Paradigm shift in urban energy systems through distributed generation: Methods and models,” *Appl. Energy*, vol. 88, no. 4, pp. 1032–1048, Apr. 2011.

[70] K. Strunz, E. Abbasi, and D. N. Huu, “DC Microgrid for Wind and Solar Power Integration,” *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 2, no. 1, pp. 115–126, Mar. 2014.

[71] L. Zhang, N. Gari, and L. V. Hmurcik, “Energy management in a microgrid with distributed energy resources,” *Energy Convers. Manag.*, vol. 78, pp. 297–305, Feb. 2014.

[72] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, “Virtual Power Plants,” in *Integrating Renewables in Electricity Markets*, vol. 205, 2014, pp. 243–287.

[73] C. Binding *et al.*, “Electric vehicle fleet integration in the danish EDISON project - A virtual power plant on the island of Bornholm,” in *Power and Energy Society General Meeting, 2010 IEEE*, 2010, pp. 1–8.

[74] S. Bracco, F. Delfino, F. Pampararo, M. Robba, and M. Rossi, “A system of systems model for the control of the university of Genoa Smart Polygeneration Microgrid,” 2012, pp. 7–12.

[75] A. Bonfiglio *et al.*, “An optimization algorithm for the operation planning of the University of Genoa smart polygeneration microgrid,” 2013, pp. 1–8.

[76] M. Rivarolo, A. Greco, and A. F. Massardo, “Thermo-economic optimization of the impact of renewable generators on poly-generation smart-grids including hot thermal storage,” *Energy Convers. Manag.*, vol. 65, no. 0, pp. 75–83, 2013.

[77] A. L. Facci, L. Andreassi, and S. Ubertini, “Optimization of CHCP (combined heat power and cooling) systems operation strategy using dynamic programming,” *Energy*, vol. 66, pp. 387–400, Mar. 2014.

[78] J.-H. Lee and H.-M. Kim, “LP-based Mathematical Model for Optimal Microgrid Operation Considering Heat Trade with District Heat System.,” *Int. J. Energy Inf. Commun.*, vol. 4, no. 4, 2013.

[79] D. Henning, S. Amiri, and K. Holmgren, “Modelling and optimisation of electricity, steam and district heating production for a local Swedish utility,” *Eur. J. Oper. Res.*, vol. 175, no. 2, pp. 1224–1247, Dec. 2006.

[80] P. Mancarella, “MES (multi-energy systems): An overview of concepts and evaluation models,” *Energy*, Nov. 2013.

[81] T. Logenthiran, D. Srinivasan, and T. Z. Shun, “Demand Side Management in Smart Grid Using Heuristic Optimization,” *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1244–1252, Sep. 2012.

[82] P. Palensky and D. Dietrich, “Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads,” *Ind. Inform. IEEE Trans. On*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[83] G. Kyriakarakos, D. D. Piromalis, A. I. Dounis, K. G. Arvanitis, and G. Papadakis, “Intelligent demand side energy management system for autonomous polygeneration microgrids,” *Appl. Energy*, vol. 103, pp. 39–51, 2013.

[84] Q. Wang, C. Zhang, Y. Ding, G. Xydis, J. Wang, and J. Østergaard, “Review of real-time electricity markets for integrating Distributed Energy Resources and Demand Response,” *Appl. Energy*, vol. 138, pp. 695–706, Jan. 2015.

[85] R. Baños, F. Manzano-Agugliaro, F. G. Montoya, C. Gil, A. Alcayde, and J. Gómez, “Optimization methods applied to renewable and sustainable energy: A review,” *Renew. Sustain. Energy Rev.*, vol. 15, no. 4, pp. 1753–1766, May 2011.

[86] A. A. Bazmi and G. Zahedi, “Sustainable energy systems: Role of optimization modeling techniques in power generation and supply: A review,” *Renew. Sustain. Energy Rev.*, vol. 15, no. 8, pp. 3480–3500, 2011.

[87] A. D. Dominguez-Garcia, C. N. Hadjicostis, and N. H. Vaidya, “Resilient Networked Control of Distributed Energy Resources,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 6, pp. 1137–1148, Jul. 2012.

[88] S. B. Ghosn, P. Ranganathan, S. Salem, J. Tang, D. Loegering, and K. E. Nygard, “Agent-oriented designs for a self healing smart grid,” in *Smart Grid*

*Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 461–466.

[89] D. Li, S. Wang, J. Zhan and Y. Zhao, "A self-healing reconfiguration technique for smart distribution networks with DGs," 2011 International Conference on Electrical and Control Engineering, Yichang, 2011, pp. 4318-4321.

[90] W. Su, J. Wang, and J. Roh, “Stochastic Energy Scheduling in Microgrids With Intermittent Renewable Energy Resources,” *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1876–1883, Jul. 2014.

[91] H. S. V. S. Kumar Nunna and S. Doolla, “Multiagent-Based Distributed-Energy-Resource Management for Intelligent Microgrids,” *Ind. Electron. IEEE Trans. On*, vol. 60, no. 4, pp. 1678–1687, Apr. 2013.

[92] K. Baclawski, “Semantic Interoperability for Big Data,” College of Computer and Information Science Northeastern University, 2014.

[93] D. Von Dollen, “Report to NIST on the smart grid interoperability standards roadmap,” *Prep. Electr. Power Res. Inst. NIST June 2009*, 2009.

[94] F. H. Abanda, J. H. M. Tah, and R. Keivani, “Trends in built environment semantic Web applications: Where are we today?,” *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5563–5577, Oct. 2013.

[95] A. Wagner, S. Speiser, and A. Harth, “Semantic Web Technologies for a Smart Energy Grid: Requirements and Challenges.,” in *ISWC Posters&Demos*, 2010.

[96] A Sicilia, L Madrazo, and J Pleguezuelos, “Integrating multiple data sources, domains and tools in urban energy models using semantic technologies,” in *eWork and eBusiness in Architecture, Engineering and Construction*, CRC Press, 2014, pp. 837–844.

[97] Q. Zhou, S. Natarajan, Y. Simmhan, and V. Prasanna, “Semantic Information Modeling for Emerging Applications in Smart Grid,” in *Information Technology: New Generations (ITNG), 2012 Ninth International Conference on*, 2012, pp. 775–782.

[98] S. Tomic, A. Fensel, and T. Pellegrini, “SESAME Demonstrator: Ontologies, Services and Policies for Energy Efficiency,” *- Semant.*, 2010.

[99] H. A. J. Schevers and R. M. Drogemuller, “Semantic Web for an Integrated Urban Software System,” in *MODSIM*, 2005, vol. 5, pp. 2040–2046.

[100] J. Teller, R. Billen, and A.-F. Cutting-Decele, “ONTOLOGY-BASED APPROACHES FOR IMPROVING THE INTEROPERABILITY BETWEEN 3D URBAN MODELS,” 2010.

[101] K. H. Van Dam and J. Keirstead, “Re-use of an ontology for modelling urban energy systems,” in *3rd International Conference on Next Generation Infrastructure Systems for Eco-Cities, INFRA 2010 - Conference Proceedings*, 2010.

[102] F. H. Abanda, J. H. M. Tah, and R. Keivani, "Trends in built environment semantic Web applications: Where are we today?," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5563–5577, 2013.

[103] Global Economic Forum, "The Global Risk Report 2016," Global Economic Forum, 2016.

[104] Impax Asset Management, "Investing in water: tapping into a source of resilient growth," Impax Asset Management, Nov. 2016.

[105] DEFRA, "Reform of the water industry: retail competition." DEFRA, Nov-2013.

[106] C. Cherchi, M. Badruzzaman, J. Oppenheimer, C. M. Bros, and J. G. Jacangelo, "Energy and water quality management systems for water utility's operations: A review," *J. Environ. Manage.*, vol. 153, pp. 108–120, Apr. 2015.

[107] Sensus, "Water 20/20 Bringing smart water networks into focus," 2012.

[108] E. Creaco, E. Lanfranchi, C. Chiesa, M. Fantozzi, C. A. Carrettini, and M. Franchini, "Optimisation of leakage and energy in the Abbiategrasso district," *Civ. Eng. Environ. Syst.*, vol. 33, no. 1, pp. 22–34, Jan. 2016.

[109] T. Boyle *et al.*, "Intelligent Metering for Urban Water: A Review," *Water*, vol. 5, no. 3, pp. 1052–1081, Jul. 2013.

[110] A. Peleg, "Investment in smart water networks on the rise," *Water Technol.*, 2014.

[111] ICT4Water, "ICT4Water standards report," European Commission, 2015.

[112] A. Hauser *et al.*, "Communication in Smart Water Networks," SWAN Forum Interoperability Workgroup, SWAN, Jun. 2016.

[113] N. Damour, "Interoperability for Water Management in Smart Cities," 2016 Smart Water Network Forum Conference, 2016.

[114] S. Harrison, "WITS-DNP3 Protocol," 2016 Smart Water Network Forum Conference, 2016.

[115] Department for Business Innovation & Skills, "SMART CITIES: background paper." 2013.

[116] R. Wenge, X. Zhang, C. Dave, L. Chao, and S. Hao, "Smart city architecture: A technology guide for implementation and design challenges," *China Commun.*, vol. 11, no. 3, pp. 56–69, 2014.

[117] R. Goodspeed, "Smart cities: moving beyond urban cybernetics to tackle wicked problems," *Camb. J. Reg. Econ. Soc.*, vol. 8, no. 1, pp. 79–92, Mar. 2015.

[118] Google, "'smart city', 'digital city', 'green city', 'future city', 'eco city' - Explore - Google Trends," 2017. [Online]. Available: https://trends.google.co.uk/trends/explore?date=all&q=%22smart%20city%2

2,%22digital%20city%22,%22green%20city%22,%22future%20city%22,%22eco%20city%22. [Accessed: 16-Mar-2017].

[119] M. Hara, T. Nagao, S. Hannoe, and J. Nakamura, "New Key Performance Indicators for a Smart Sustainable City," *Sustainability*, vol. 8, no. 3, p. 206, Mar. 2016.

[120] M. Krarti, "An Overview of Artificial Intelligence-Based Methods for Building Energy Systems," *J. Sol. Energy Eng.*, vol. 125, no. 3, p. 331, 2003.

[121] J. K. W. Wong, H. Li, and S. W. Wang, "Intelligent building research: a review," *Autom. Constr.*, vol. 14, no. 1, pp. 143–159, Jan. 2005.

[122] Z. Ma and S. Wang, "Building energy research in Hong Kong: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 8, pp. 1870–1883, Oct. 2009.

[123] A. I. Dounis and C. Caraiscos, "Advanced control systems engineering for energy and comfort management in a building environment—A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 6–7, pp. 1246–1261, Aug. 2009.

[124] K. Marik, J. Rojicek, P. Stluka, and J. Vass, "Advanced HVAC Control: Theory vs Reality," *IFAC World Congr.*, 2011.

[125] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in Smart City initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, Jun. 2014.

[126] D.-H. Shin, "Ubiquitous city: Urban technologies, urban infrastructure and urban informatics," *J. Inf. Sci.*, vol. 35, no. 5, pp. 515–526, Oct. 2009.

[127] F. Gil-Castineira, E. Costa-Montenegro, F. J. Gonzalez-Castano, C. López-Bravo, T. Ojala, and R. Bose, "Experiences inside the ubiquitous oulu smart city," *Computer*, no. 6, pp. 48–55, 2011.

[128] A. Mahizhnan, "Smart cities: the Singapore case," *Cities*, vol. 16, no. 1, pp. 13–18, 1999.

[129] E. Moir, T. Moonen, and G. Clark, "WHAT ARE FUTURE CITIES? ORIGINS, MEANINGS AND USES." Future Cities Catapult, 2014.

[130] BSI, "PAS181:2014." 2014.

[131] BSI, "PD8101-2014." 2014.

[132] BSI, "Smart Cities - Vocabulary," BSI, PAS 180:2014, 2014.

[133] Andries van Dijk *et al.*, "Smart Cities: How rapid advances in technology are reshaping our economy and society," Deloitte, The Netherlands, Nov. 2015.

[134] Sarwant Singh, "Smart Cities -- A $1.5 Trillion Market Opportunity," 2014. [Online]. Available: https://www.forbes.com/sites/sarwantsingh/2014/06/19/smart-cities-a-1-5-trillion-market-opportunity/#16535b5c6053. [Accessed: 17-Mar-2017].

[135] Holli Haswell, "IBM News room - 2013-03-31 Frost & Sullivan Leadership Award Names IBM Best-in-Class Global Smart City Integrator - United

States," 2013. [Online]. Available: http://www-03.ibm.com/press/us/en/pressrelease/43562.wss#release. [Accessed: 17-Mar-2017].

[136] Y. Rezgui, "Cardiff Future Cities Final Stage Application," Cardiff University, 2012. [Online]. Available: https://connect.innovateuk.org/documents/3130726/3794125/Feasibility+Study+-+Cardiff+City+Council.pdf/2f29edd1-7d83-4b5e-b675-819e87cae4ed. [Accessed 26-May-17]

[137] ARUP, "Solutions for Cities - An analysis of the Feasibility Studies from the Future Cities Demonstrator Programme." 2013.

[138] S. Zygiaris, "Smart City Reference Model: Assisting Planners to Conceptualize the Building of Smart City Innovation Ecosystems," *J. Knowl. Econ.*, vol. 4, no. 2, pp. 217–231, Jun. 2013.

[139] I. Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, and X. Vilajosana, "Bootstrapping smart cities through a self-sustainable model based on big data flows," *Commun. Mag. IEEE*, vol. 51, no. 6, pp. 128–134, 2013.

[140] G. S. Yovanof and G. N. Hazapis, "An Architectural Framework and Enabling Wireless Technologies for Digital Cities & Intelligent Urban Environments," *Wirel. Pers. Commun.*, vol. 49, no. 3, pp. 445–463, May 2009.

[141] C. Harrison and I. A. Donnelly, "A theory of smart cities," in *Proceedings of the 55th Annual Meeting of the ISSS-2011, Hull, UK*, 2011, vol. 55.

[142] N. Komninos, "What makes cities Smart?," in *presentation, Smart Cities Conference, Edinburgh*, 2011, vol. 30.

[143] Arup (last) and Technology Strategy Board (last), "Solutions for Cities-An analysis of the feasibility studies from the Future Cities Demonstrator Programme," 2013.

[144] Glasgow City Council, "Future City Glasgow," 2015. [Online]. Available: http://futurecity.glasgow.gov.uk/. [Accessed: 26-May-2015].

[145] S. Revill, "A New Data Landscape: Delivering Open Data for the City of Glasgow - …," 2014.

[146] Future City Glasgow, "Glasgow Operations Centre," 2015. [Online]. Available: http://futurecity.glasgow.gov.uk/ops-data/. [Accessed: 27-May-2015].

[147] Future City Glasgow, "Open Data," 2015. [Online]. Available: http://futurecity.glasgow.gov.uk/data/. [Accessed: 26-May-2015].

[148] Future City Glasgow, "OPEN Glasgow | Linked data," 2014. [Online]. Available: http://open.glasgow.gov.uk/linked-data-glasgow/. [Accessed: 27-May-2015].

[149] Step Up Smart Cities Project, "Glasgow Technology Strategy Board." 2013.

[150] Future City Glasgow, “Asset Mapping,” 2014. [Online]. Available: http://open.glasgow.gov.uk/load/. [Accessed: 27-May-2015].

[151] City of London, “Smart London vision,” 2016. [Online]. Available: http://www.london.gov.uk/priorities/business-economy/vision-and-strategy/smart-london. [Accessed: 03-Jun-2015].

[152] IMechE, “Top ten smart cities in the world,” 2014. [Online]. Available: http://www.imeche.org/news/blog/top-ten-smart-cities-in-the-world. [Accessed: 28-May-2015].

[153] A.-V. Anttiroiko, P. Valkama, and S. J. Bailey, “Smart cities in the new service economy: building platforms for smart services,” *AI Soc.*, vol. 29, no. 3, pp. 323–334, Aug. 2014.

[154] A. K. Debnath, H. C. Chin, M. M. Haque, and B. Yuen, “A methodological framework for benchmarking smart transport cities,” *Cities*, vol. 37, pp. 47–56, Apr. 2014.

[155] Mayor of London, “London Heat Network manual.” Greater London Authority, 2014.

[156] London City Hall, “Londoners at the core,” 2014. [Online]. Available: http://www.london.gov.uk/priorities/business-economy/vision-and-strategy/smart-london/londoners-at-the-core. [Accessed: 03-Jun-2015].

[157] “Access to open data | London City Hall,” 2015. [Online]. Available: http://www.london.gov.uk/priorities/business-economy/vision-and-strategy/smart-london/access-to-open-data. [Accessed: 03-Jun-2015].

[158] Greater London Authority, “London Datastore,” 2015. [Online]. Available: http://data.london.gov.uk/. [Accessed: 03-Jun-2015].

[159] Centre for Cities, “Smart Cities briefing,” 2014.

[160] P. High, “The Top Five Smart Cities In The World - Forbes,” 2015. [Online]. Available: http://www.forbes.com/sites/peterhigh/2015/03/09/the-top-five-smart-cities-in-the-world/. [Accessed: 28-May-2015].

[161] “Barcelona Named ‘Global Smart City – 2015’ - Juniper Research.” [Online]. Available: http://www.juniperresearch.com/press/press-releases/barcelona-named-global-smart-city-2015. [Accessed: 28-May-2015].

[162] City Climate Leadership Awards, “Barcelona: Barcelona Smart City,” 2013. [Online]. Available: http://cityclimateleadershipawards.com/barcelona-barcelona-smart-city/. [Accessed: 28-May-2015].

[163] B. Cohen, “Barcelona: A Smart City Model For The Planet,” *www.fastcodesign.com*, 2011. [Online]. Available: http://www.fastcoexist.com/1679017/barcelona-a-smart-city-model-for-the-planet. [Accessed: 28-May-2015].

[164] T. Sheu, “Smart City Barcelona: One City’s Quest For Innovation,” *www.barcinno.com*, 2015. [Online]. Available: http://www.barcinno.com/smart-city-barcelona/. [Accessed: 28-May-2015].

[165] J. Ancheta, “Ten Reasons Why Barcelona is a Smart City - VilaWeb,” *VilaWeb*, 2014. [Online]. Available: http://www.vilaweb.cat/noticia/4175829/20140226/ten-reasons-why-barcelona-is-smart-city.html. [Accessed: 28-May-2015].

[166] Department for Business Innovation & Skills, “Global Innovators: International Case Studies on Smart Cities,” 135, 2013.

[167] T. Bakıcı, E. Almirall, and J. Wareham, “A Smart City Initiative: the Case of Barcelona,” *J. Knowl. Econ.*, vol. 4, no. 2, pp. 135–148, Jun. 2013.

[168] BCN Smart City, “Smart City Areas,” 2015. [Online]. Available: http://smartcity.bcn.cat/en/smart-city-areas.html. [Accessed: 31-May-2015].

[169] BCN Smart City, “City OS,” 2015. [Online]. Available: http://smartcity.bcn.cat/en/city-os.html. [Accessed: 31-May-2015].

[170] Ajuntament de Barcelona, “Inside the City OS,” 2012.

[171] J. M. Cela, “BSC Smart Cities Initiative,” 2014.

[172] Katia Moskvitch, “Barcelona: the world’s smartest city?,” E&T Magazine, 2016.

[173] Smart Nation Programme Office, “Singapore: Smart Nation,” 2016. [Online]. Available: http://www.smartnation.sg/. [Accessed: 04-Nov-2016].

[174] J. Watts and N. Purnell, “Singapore Is Taking the ‘Smart City’ to a Whole New Level,” *Wall Street Journal*, 2016.

[175] L. Sun and K. W. Axhausen, “Understanding urban mobility patterns with a probabilistic tensor factorization framework,” *Transp. Res. Part B Methodol.*, vol. 91, pp. 511–524, Sep. 2016.

[176] A. I. Niculescu *et al.*, “Designing IDA - An Intelligent Driver Assistant for Smart City Parking in Singapore,” in *Human-Computer Interaction – INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part IV*, J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds. Cham: Springer International Publishing, 2015, pp. 510–513.

[177] C. E. Seng, “Singapore’s smart nation program—Enablers and challenges,” in *System of Systems Engineering Conference (SoSE), 2016 11th*, 2016, pp. 1–5.

[178] P. Berrone and J. E. Ricart, “IESE Insight New York Edges Out London as the World’s ‘Smartest’ City,” IESE, 2016.

[179] White House, “FACT SHEET: Announcing Over $80 million in New Federal Investment and a Doubling of Participating Communities in the White House Smart Cities Initiative | whitehouse.gov,” 2016.

[180] J. Frazier and T. Touchet, “Transforming the City of New York New: Platform for Public-Private Cooperation Ushers in Smart Cities of the Future,” Cisco, 2012.

[181] A. Greenfield, *Against the smart city*. Do projects, 2013.

[182] V. Niaros, “Introducing a Taxonomy of the ‘Smart City’: Towards a Commons-Oriented Approach?,” *TripleC Commun. Capital. Crit. Open Access J. Glob. Sustain. Inf. Soc.*, vol. 14, no. 1, pp. 51–61, 2016.

[183] T. Saunders and P. Baeck, “Rethinking Smart Cities from the Ground Up,” *Lond. Nesta*, 2015.

[184] C. Hastie, “Benefits of Urban Trees,” Warwick District Council, 2003.

[185] F. S. Ferraz and C. A. G. Ferraz, “More than Meets the Eye in Smart City Information Security: Exploring Security Issues Far beyond Privacy Concerns,” 2014, pp. 677–685.

[186] D. Dhungana, G. Engelbrecht, J. X. Parreira, A. Schuster, and D. Valerio, “Aspern smart ICT: Data analytics and privacy challenges in a smart city,” in *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, 2015, pp. 447–452.

[187] K. E. Skouby, A. Kivimäki, L. Haukiputo, P. Lynggaard, and I. M. Windekilde, “Smart Cities and the Ageing Population,” in *The 32nd Meeting of WWRF*, 2014.

[188] A. Degbelo, C. Granell, S. Trilles, D. Bhattacharya, S. Casteleyn, and C. Kray, “Opening up Smart Cities: Citizen-Centric Challenges and Opportunities from GIScience,” *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 2, p. 16, Feb. 2016.

[189] Stop Smart Meters, “Health Effects,” 2012. .

[190] C. Sage and D. O. Carpenter, “Public health implications of wireless technologies,” *Pathophysiology*, vol. 16, no. 2–3, pp. 233–246, Aug. 2009.

[191] K. R. Foster and J. E. Moulder, “Wi-Fi and health: Review of current status of research,” *Health Phys.*, vol. 105, no. 6, pp. 561–575, 2013.

[192] M. Martínez-Burdalo, A. Martín, A. Sanchis, and R. Villar, “FDTD assessment of human exposure to electromagnetic fields from WiFi and bluetooth devices in some operating situations,” *Bioelectromagnetics*, vol. 30, no. 2, pp. 142–151, 2009.

[193] K. R. Foster, “Radiofrequency exposure from wireless lans utilizing Wi-Fi technology,” *Health Phys.*, vol. 92, no. 3, pp. 280–289, 2007.

[194] E. Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*. New York: Oxford University Press, 2008.

[195] Information Commissioner’s Office, “Big data and data protection,” 2014.

[196] Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights,” The White House, 2016.

[197] J. Polonetsky, O. Tene, and J. Jerome, “Benefit-Risk Analysis for Big Data Projects,” in *Future of Privacy Forum (FPF). Washington, DC*, 2014.

[198] T. Peltan, “Smart Cities as complexity management,” in *Smart Cities Symposium Prague (SCSP), 2015*, 2015, pp. 1–5.

[199] L. von Bertalanffy, *General System Theory*. New York: George Braziller, 1968.

[200] B. P. Thompson and L. C. Bank, “Use of system dynamics as a decision-making tool in building design and operation,” *Build. Environ.*, vol. 45, no. 4, pp. 1006–1015, 2010.

[201] Moonseo Park, Sae-Hyun Ji, Hyun-Soo Lee, and Wooyoung Kim, “Strategies for Design-Build in Korea Using System Dynamics Modeling,” *J. Constr. Eng. Manag.*, vol. 135, pp. 1125–1137, 2009.

[202] N. Suh, *Complexity: Theory and Applications*. Oxford: Oxford University Press, 2005.

[203] J Hilton and C Wright, “Building Resilience into Systems,” presented at the IEEE Systems Conference, 2012.

[204] J. Fiksel, “Designing Resilient, Sustainable Systems,” *Evironmental Sci. Technol.*, vol. 37, 2003.

[205] A. Urken, “Designing evolvable systems in a framework of robust, resilient and sustainable engineering analysis,” *Adv. Eng. Inform.*, vol. 26, 2012.

[206] C. Wright, V. Kiparoglou, M. Williams, and J. Hilton, “A Framework for Resilience Thinking,” *Procedia Comput. Sci.*, vol. 8, pp. 45–52, 2012.

[207] V. Vlad, C. Buzduga, and C. Ciufudean, “An Approach to Developing Power Grid Control Systems with IEC 61850, IEC 61499 and Holonic Control,” *WSEAS Trans. Syst.*, vol. 13, pp. 503–509, 2014.

[208] A. Koestler, *Beyond Reductionism: New Perspectives in the Life Sciences*. London: Hutchinson, 1969.

[209] B. Roy, “Using Agents to Make and Manage Markets Across a Supply Web,” *Complexity*, vol. 3, no. 4, 1998.

[210] T. Ahram, W. Karwowski, B. Amaba, and Fechtelkotter, *Power and Energy Management: A User-Centered System-of-Systems Engineering Approach*. Hoboken, N.J: John Wiley & Sons, 2013.

[211] V. Javidroozi, H. Shah, A. Cole, and A. Amini, “Towards a City’s Systems Integration Model for Smart City Development: A Conceptualization,” 2015, pp. 312–317.

[212] Y. Rezgui, “Exploring virtual team-working effectiveness in the construction sector,” *Interact. Comput.*, vol. 19, no. 1, pp. 96–112, Jan. 2007.

[213] G. Rubio, J. Martínez, D. Gómez, and X. Li, “Semantic Registration and Discovery System of Subsystems and Services within an Interoperable Coordination Platform in Smart Cities,” *Sensors*, vol. 16, no. 7, p. 955, Jun. 2016.

[214] G. Fuchs, “Beyond the Dashboard,” Spotfire, Inc., Somerville, MA, 2005.

[215] UCL, “CityDashboard: London,” 2012. [Online]. Available: http://citydashboard.org/london/. [Accessed: 30-Apr-2016].

[216] A. Singleton and O. O’Brien, “DataShine: 2011 OAC,” *DataShine: 2011 OAC*, 2011. [Online]. Available: http://oac.datashine.org.uk/. [Accessed: 30-Apr-2016].

[217] Freeboard, “Dashboards For the Internet Of Things,” 2014. [Online]. Available: https://freeboard.io/. [Accessed: 30-Apr-2016].

[218] A. Morales-Torres, I. Escuder-Bueno, I. Andrés-Doménech, and S. Perales-Momparler, “Decision Support Tool for energy-efficient, sustainable and integrated urban stormwater management,” *Environ. Model. Softw.*, vol. 84, pp. 518–528, Oct. 2016.

[219] A. Kaklauskas, “Intelligent Decision Support Systems,” in *Biometric and Intelligent Decision Making Support*, vol. 81, Cham: Springer International Publishing, 2015, pp. 31–85.

[220] J. W. Moon, S. K. Jung, Y. Kim, and S.-H. Han, “Comparative study of artificial intelligence-based building thermal control methods – Application of fuzzy, adaptive neuro-fuzzy inference system, and artificial neural network,” *Appl. Therm. Eng.*, vol. 31, no. 14–15, pp. 2422–2429, Oct. 2011.

[221] B. Yuce, Y. Rezgui, and S. Howell, “ANN-genetic algorithm-based rule generation for holistic energy management in public buildings,” in *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling*, 2015.

[222] G. Weiss, *Multiagent systems a modern approach to distributed artificial intelligence*. Cambridge, Mass.: MIT Press, 1999.

[223] B. Hosack, D. Hall, D. Paradice, and J. F. Courtney, “A look toward the future: decision support systems research is alive and well,” *J. Assoc. Inf. Syst.*, vol. 13, no. 5, p. 315, 2012.

[224] A. Voinou, “Data Lakes – where do they fit in the Data landscape?,” Atlis, 2016.

[225] S. Katta and S. Anisingaraju, “Analytics Elements to Establish Semantic Consistency in Data Lakes,” Saama, Campbell, CA, 2015.

[226] B. Stein and A. Morrison, “The enterprise data lake: Better integration and deeper analytics,” PWC, 2014.

[227] S. L. Pfleeger and J. M. Atlee, “Capturing the Requirements,” in *Software Engineering: Theory and Practice*, Pearson, 2009.

[228] A. Hefnawy, A. Bouras, and C. Cherifi, “IoT for Smart City Services: Lifecycle Approach,” 2016, pp. 1–9.

[229] Z. Irani *et al.*, “A Decision Support System for Fostering Smart Energy Efficient Districts,” 2015.

[230] S. Liu, A. H. B. Duffy, R. I. Whitfield, and I. M. Boyle, “Integration of decision support systems to improve decision support performance,” *Knowl. Inf. Syst.*, vol. 22, no. 3, pp. 261–286, Mar. 2010.

[231] E. Blomqvist, “The use of Semantic Web technologies for decision support–a survey,” *Semantic Web*, vol. 5, no. 3, pp. 177–201, 2014.

[232] A. Sheth, P. Anantharam, and C. Henson, “Semantic, Cognitive, and Perceptual Computing: Paradigms That Shape Human Experience,” *IEEE Comput. Soc.*, 2016.

[233] B. Williamson, “Computing brains: learning algorithms and neurocomputation in the smart city,” *Inf. Commun. Soc.*, vol. 20, no. 1, pp. 81–99, Jan. 2017.

[234] P. Kaltenrieder, E. Portmann, and T. Myrach, “Fuzzy knowledge representation in cognitive cities,” in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, 2015, pp. 1–8.

[235] J. F. De Paz, J. Bajo, S. Rodríguez, G. Villarrubia, and J. M. Corchado, “Intelligent system for lighting control in smart cities,” *Inf. Sci.*, vol. 372, pp. 241–255, Dec. 2016.

[236] M. G. Kienzle, “Cognitive Technologies for Smarter Cities,” 2016, pp. 12–21.

[237] W. Cabrera, D. Benhaddou, and C. Ordonez, “Solar Power Prediction for Smart Community Microgrid,” in *Smart Computing (SMARTCOMP), 2016 IEEE International Conference on*, 2016, pp. 1–6.

[238] A. Kantamneni, L. E. Brown, G. Parker, and W. W. Weaver, “Survey of multi-agent systems for microgrid control,” *Eng. Appl. Artif. Intell.*, vol. 45, pp. 192–203, Oct. 2015.

[239] KES-AMSTA 2009, Ed., *Agent and multi-agent systems*. Berlin; New York: Springer, 2009.

[240] J. M. Guerrero, M. Chandorkar, T.-L. Lee, and P. C. Loh, “Advanced Control Architectures for Intelligent Microgrids—Part I: Decentralized and Hierarchical Control,” *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1254–1262, Apr. 2013.

[241] P. Booij, V. Kamphuis, O. van Pruissen, and C. Warmer, “Multi-agent control for integrated heat and electricity management in residential districts,” *ATES AAMAS*, vol. 13, pp. 6–10, 2013.

[242] T. Linnenberg, I. Wior, S. Schreiber, and A. Fay, “A market-based multi-agent-system for decentralized power and grid control,” in *Emerging Technologies & Factory Automation (ETFA), 2011 IEEE 16th Conference on*, 2011, pp. 1–8.

[243] Z. Yang, C. Ma, J. Q. Feng, Q. H. Wu, S. Mann, and J. Fitch, “A multi-agent framework for power system automation,” *Int. J. Innov. Energy Syst. Power*, vol. 1, no. 1, pp. 39–45, 2006.

[244] S. Mullen and G. Onsongo, “Decentralized agent-based underfrequency load shedding,” *Integr. Comput.-Aided Eng.*, vol. 17, no. 4, pp. 321–329, Jan. 2010.

[245] J. M. Solanki, S. K. Solanki, and N. Schulz, “Multi-agent-based reconfiguration for restoration of distribution systems with distributed generators,” *Integr. Comput.-Aided Eng.*, vol. 17, no. 4, pp. 331–346, Jan. 2010.

[246] M. L Brodie, “Future Intelligent Information Systems: AI and Database Technologies Working Together,” presented at the AAAI-88, 1988.

[247] M Brodie, M Greaves, and J Hendler, “Databases and AI: The Twain Just Met,” presented at the STI Semantic Summit, 2011.

[248] BuildingSMART, “IFC4 Add2.” Jul-2016.

[249] SEMANCO, “Deliverable 4.2 Semantic Energy Model.” 2013.

[250] A. Crapo, K. Griffith, A. Khandelwal, J. Lizzi, A. Moitra, and X. Wang, “Overcoming challenges using the CIM as a semantic model for energy applications,” in *Grid-Interop Forum, GridWise Architecture Council*, 2010.

[251] WatERP, “D1.3 Generic ontology for water supply distribution chain.” 2013.

[252] M. Bermudez-Edo, T. Elsaleh, P. Barnaghi, and K. Taylor, “IoT-Lite Ontology,” 2015. [Online]. Available: https://www.w3.org/Submission/2015/SUBM-iot-lite-20151126/. [Accessed: 11-Nov-2016].

[253] JL Dessalles, JP Muller, and D Phan, “Emergence in multi-agent systems: conceptual and methodological issues,” in *MultiAgent Models and Simulation for Social and Human Sciences*, Oxford: The Bardwell-Press, 2007.

[254] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Netw.*, vol. 10, no. 7, pp. 1497–1516, Sep. 2012.

[255] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[256] L. Atzori, A. Iera, and G. Morabito, “The Internet of Things: A survey,” *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[257] A. W. Burange and H. D. Misalkar, “Review of Internet of Things in development of smart cities with data management & privacy,” in *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in*, 2015, pp. 189–195.

[258] AIOTI, “Internet of Things Applications AIOTI WG01 – IERC,” 2015.

[259] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, “Smart community: an internet of things application,” *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 68–75, 2011.

[260] A. J. Jara, M. A. Zamora, and A. F. G. Skarmeta, “An internet of things–based personal device for diabetes therapy management in ambient assisted living (AAL),” *Pers. Ubiquitous Comput.*, vol. 15, no. 4, pp. 431–440, Apr. 2011.

[261] S. Karnouskos, “Cyber-Physical Systems in the SmartGrid,” 2011, pp. 20–23.

[262] N. Bui, A. P. Castellani, P. Casari, and M. Zorzi, “The internet of energy: a web-enabled smart grid system,” *IEEE Netw.*, vol. 26, no. 4, pp. 39–45, 2012.

[263] M. Yun and B. Yuxin, “Research on the architecture and key technology of Internet of Things (IoT) applied on smart grid,” in *Advances in Energy Engineering (ICAEE), 2010 International Conference on*, 2010, pp. 69–72.

[264] L. H. Tsoukalas and R. Gao, “From smart grids to an energy internet: Assumptions, architectures and requirements,” in *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, 2008, pp. 94–98.

[265] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, “A knowledge-based approach for real-time IoT data stream annotation and processing,” in *Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE*, 2014, pp. 215–222.

[266] ALMANAC, “D6.1 A Scalable Data Management Architecture for Smart City Environments.” 2014.

[267] RERUM, “RERUM deliverable D2.3: System Architecture.” 2014.

[268] R. Lea and M. Blackstock, “City Hub: A Cloud-Based IoT Platform for Smart Cities,” 2014, pp. 799–804.

[269] CityPulse, “D3.1:Semantic Data Stream Annotation for Automated Framework.” 2014.

[270] ALMANAC, “D8.2 Application Definition - Water Management.” 2014.

[271] VITAL, “D2.3: Virtualization Architecture and Technical Specifications.” 2015.

[272] TRESCIMO, “D3.1: Architecture Specification.” 2014.

[273] ClouT, “D3.2 -City Platform as a Service (CPaaS).” 2014.

[274] SMARTIE, “Initial Architecture Specification.” 2014.

[275] FIESTA, “Analysis of IoT Platforms and Testbeds.” 2015.

[276] CISCO, “IoT Reference Model,” Jun-2014.

[277] W3C, “WoT Current Practices,” 2016. [Online]. Available: http://w3c.github.io/wot/current-practices/wot-practices.html. [Accessed: 11-Nov-2016].

[278] OneM2M, “TR-0025 Application Developer Guide.” OneM2M, Feb-2016.

[279] OneM2M, “TS-0004 Service Layer Core Protocol.” OneM2M, 2016.

[280] OneM2M, “TS-0001 Functional Architecture.” OneM2M, 2016.

[281] OGC, “SensorThings API - Part 1: Sensing,” 2015. [Online]. Available: http://docs.opengeospatial.org/is/15-078r6/15-078r6.html. [Accessed: 05-Oct-2016].

[282] P. Waher and R. Klauck, “XEP-0347: Internet of Things - Discovery,” *XMPP Standards Foundation*, 2016. [Online]. Available: https://xmpp.org/extensions/xep-0347.html. [Accessed: 22-Mar-2017].

[283] P Gayathri, H Santhi, N. Ashish, and G. Shubham, “Study and analysis of privacy and security challenges in internet of things (IOT),” *Int. J. Pharm. Technol.*, vol. 8, no. 3, 2016.

[284] A. Benin, S. Toledo, and E. Tromer, “Secure Association for the Internet of Things,” in *2015 International Workshop on Secure Internet of Things (SIoT)*, 2015, pp. 25–34.

[285] J. R. Nurse, A. Erola, I. Agrafiotis, M. Goldsmith, and S. Creese, “Smart Insiders: Exploring the Threat from Insiders using the Internet-of-Things,” in *2015 International Workshop on Secure Internet of Things (SIoT)*, 2015, pp. 5–14.

[286] P. Fremantle, B. Aziz, J. Kopecky, and P. Scott, “Federated Identity and Access Management for the Internet of Things,” 2014, pp. 10–17.

[287] A. W. Atamli and A. Martin, “Threat-Based Security Analysis for the Internet of Things,” 2014, pp. 35–43.

[288] M. Abomhara and G. M. Køien, “Security and privacy in the Internet of Things: Current status and open issues,” in *Privacy and Security in Mobile Systems (PRISMS), 2014 International Conference on*, 2014, pp. 1–8.

[289] Y. H. Hwang, “IoT Security & Privacy: Threats and Challenges,” 2015, pp. 1–1.

[290] J. H. Ziegeldorf, O. G. Morchon, and K. Wehrle, “Privacy in the Internet of Things: threats and challenges: Privacy in the Internet of Things: threats and challenges,” *Secur. Commun. Netw.*, vol. 7, no. 12, pp. 2728–2742, Dec. 2014.

[291] E. Vasilomanolakis, J. Daubert, M. Luthra, V. Gazis, A. Wiesmaier, and P. Kikiras, “On the Security and Privacy of Internet of Things Architectures and Systems,” in *2015 International Workshop on Secure Internet of Things (SIoT)*, 2015, pp. 49–57.

[292] W. W. Ren, L. Hu, K. Zhao, and B. Jia, “Information Security Ontology Model for Internet of Thing,” *Adv. Mater. Res.*, vol. 694–697, pp. 2466–2470, 2013.

[293] Humayun Zafar, “Cybersecurity and its 10 domains.” University System of Georgia, 2015.

[294] M. Paul, “The Ten Best Practices for Secure Software Development,” (ISC)2, 2009.

[295] O Garcia-Morchon, S Kumar, and M Sethi, “Security Considerations in the IP-based Internet of Things,” IETF, 2016.

[296] Oracle, Wind River, “Extending Enterprise to the Edge,” 2016.

[297] M. Fahrion, “Evolving from SCADA to IoT,” 2014.

[298] S. Khoshafian, “Proven approaches for Legacy Systems Modernization,” 2011.

[299] K. Gaiser, V. Harrington, G. Loughrin, S. Meyer, and J. Sartini, “Integrating IoT Sensor Technology into the Enterprise.” Intel, 2015.

[300] M. Zemlicka and J. Kral, “Legacy Systems and Web Services,” Charles University, Prague, 2004.

[301] M. Yoshioka, T. Sudo, A. Yoshikawa, and K. Sakata, “Legacy system integration technology for legacy application utilization from distributed object environment,” *Hitachi Rev.*, vol. 47, no. 6, pp. 284–290, 1998.

[302] B. Glimm and H. Stuckenschmidt, “15 Years of Semantic Web,” *Künstl. Intell.*, vol. 30, no. 2, pp. 117–130, 2016.

[303] P. Andersen, *What is Web 2.0?: ideas, technologies and implications for education*, vol. 1. JISC Bristol, 2007.

[304] I. Herman, “Semantic Web Adoption and Applications,” 2012.

[305] J. Cardiff, “The Evolution of the Semantic Web,” Puebla, Mexico, 2013.

[306] S. Ismail and T. Shaikh, “A Literature Review on Semantic Web - Understanding the Pioneers’ Perspective,” 2016, pp. 15–28.

[307] A. Gyrard, C. Bonnet, K. Boudaoud, and M. Serrano, “Assisting IoT Projects and Developers in Designing Interoperable Semantic Web of Things Applications,” 2015, pp. 659–666.

[308] S. Ali, S. Khusro, and H.-T. Chang, “POEM: Practical ontology engineering model for semantic web ontologies,” *Cogent Eng.*, vol. 3, no. 1, p. 1193959, Dec. 2016.

[309] J.-P. Calbimonte, S. Sarni, J. Eberle, and K. Aberer, “XGSN: an open-source semantic sensing middleware for the web of things,” in *Joint Proceedings of the 6th International Workshop on the Foundations, Technologies and Applications of the Geospatial Web, TC*, 2014, pp. 51–66.

[310] H. H. Zhu and S. Madnick, “A lightweight ontology approach to scalable interoperability,” 2006.

[311] W3C, “OWL Web Ontology Language Guide,” 2004. [Online]. Available: https://www.w3.org/TR/owl-guide/. [Accessed: 21-May-2016].

[312] M. Fernandez, A. Gomez-Perez, and N. Juristo, “METHONTOLOGY: From Ontological Art Towards Ontological Engineering,” Universidad Politecnica de Madrid, 1997.

[313] A. G. Pérez, M. C. S. de Figueroa Baonza, and B. Villazón, “Neon methodology for building ontology networks: Ontology specification,” *Methodology*, pp. 1–18, 2008.

[314] W3C, “OWL 2 Web Ontology Language Document Overview (Second Edition),” 2012. [Online]. Available: https://www.w3.org/TR/owl2-overview/. [Accessed: 22-Mar-2017].

[315] B. Liu, K. Huang, J. Li, and M. Zhou, “An Incremental and Distributed Inference Method for Large-Scale Ontologies Based on MapReduce Paradigm,” *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 53–64, Jan. 2015.

[316] W. Wang, S. De, R. Toenjes, E. Reetz, and K. Moessner, “A comprehensive ontology for knowledge representation in the internet of things,” in *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, 2012, pp. 1793–1798.

[317] S. Hachem, T. Teixeira, and V. Issarny, “Ontologies for the internet of things,” in *Proceedings of the 8th Middleware Doctoral Symposium*, 2011, p. 3.

[318] M. Ma, P. Wang, and C.-H. Chu, “Ontology-Based Semantic Modeling and Evaluation for Internet of Things Applications,” 2014, pp. 24–30.

[319] X. Wang, H. An, Y. Xu, and S. Wang, “Sensing Network Element Ontology Description Model for Internet of Things,” 2015, pp. 471–475.

[320] A. Gyrard, S. K. Datta, C. Bonnet, and K. Boudaoud, “Standardizing generic cross-domain applications in Internet of Things,” in *2014 IEEE Globecom Workshops (GC Wkshps)*, 2014, pp. 589–594.

[321] P. Edmore Chindenga, Caroline Gurajena, and Mamello Thinyane, “Towards an Adaptive Ontology Based Model for Interoperability in Internet of Things (IoT),” South Africa, 2016.

[322] Vlad Trifa, Dominique Guinard, and David Carrera, “Web Thing Model,” 2015. [Online]. Available: http://model.webofthings.io/. [Accessed: 11-Nov-2016].

[323] A. Gyrard and M. Serrano, “A Unified Semantic Engine for Internet of Things and Smart Cities: From Sensor Data to End-Users Applications,” 2015, pp. 718–725.

[324] A. Katasonov, O. Kaykova, O. Khriyenko, S. Nikitin, and V. Y. Terziyan, “Smart Semantic Middleware for the Internet of Things.,” *ICINCO-ICSO*, vol. 8, pp. 169–178, 2008.

[325] R. Mietz, S. Groppe, K. Römer, and D. Pfisterer, “Semantic Models for Scalable Search in the Internet of Things,” *J. Sens. Actuator Netw.*, vol. 2, no. 2, pp. 172–195, Mar. 2013.

[326] A. J. Jara, A. C. Olivieri, Y. Bocchi, M. Jung, W. Kastner, and A. F. Skarmeta, “Semantic web of things: an analysis of the application semantics for the IoT moving towards the IoT convergence,” *Int. J. Web Grid Serv.*, vol. 10, no. 2–3, pp. 244–272, 2014.

[327] X. Su, H. Zhang, J. Riekki, A. Keränen, J. K. Nurminen, and L. Du, “Connecting IoT Sensors to Knowledge-based Systems by Transforming SenML to RDF,” *Procedia Comput. Sci.*, vol. 32, pp. 215–222, 2014.

[328] K. Sahlmann and T. Schwotzer, “MOCAP: Towards the Semantic Web of Things,” 2015.

[329] H. Hasemann, A. Kröller, and M. Pagel, “RDF Provisioning for the Internet of Things,” in *Internet of Things (IOT), 2012 3rd International Conference on the*, 2012, pp. 143–150.

[330] M. Compton *et al.*, “The SSN ontology of the W3C semantic sensor network incubator group,” *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 17, pp. 25–32, Dec. 2012.

[331] R. A. Stewart, R. Willis, D. Giurco, K. Panuwatwanich, and G. Capati, “Web-based knowledge management system: linking smart metering to the future of urban water planning,” *Aust. Plan.*, vol. 47, no. 2, pp. 66–74, Jun. 2010.

[332] P. Barnaghi, “Semantic Technologies for the Internet of Things.” SenZations Summer School, Croatia, Sep-2014.

[333] D. Evans, “The Internet of Things, How the Next Evolution of the Inernet Is Changing Everything,” Cisco, 2011.

[334] D. Koo, K. Piratla, and C. J. Matthews, “Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (IoT),” *Procedia Eng.*, vol. 118, pp. 489–497, 2015.

[335] D. Kolb, *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs: Prentice Hall, 1984.

[336] O. Mayr, *Origins of Feedback Control*. Clinton, MA: The Colonial Press, 1970.

[337] R. P. Deb Nath, K. Hose, and T. B. Pedersen, “Towards a Programmable Semantic Extract-Transform-Load Framework for Semantic Data Warehouses,” 2015, pp. 15–24.

[338] J. M. Álvarez and A. C. López, “Integration and interoperability on service oriented architectures using semantics,” in *Proceedings of the doctoral consortium of the international conference on web engineering (ICWE) in conjunction with ICWE*, 2009, vol. 484, pp. 145–160.

[339] M. A. Bornea *et al.*, “Building an efficient RDF store over a relational database,” in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 2013, pp. 121–132.

[340] C. Aswamenakul, M. Buranarach, and K. R. Saikaew, “A review and design of framework for storing and querying rdf data using nosql database,” in *4th Joint International Semantic Technology Conference*, 2014, pp. 144–147.

[341] K. R. Saikaew, C. Asawamenakul, and M. Buranarach, “Design and evaluation of a NoSQL database for storing and querying RDF data,” *KKU Eng. J.*, vol. 41, no. 4, pp. 537–545, 2014.

[342] M. Rodriguez-Muro, M. Rezk, J. Hardi, M. Slusnys, T. Bagosi, and D. Calvanese, “Evaluating SPARQL-to-SQL translation,” in *2nd OWL Reasoner Evaluation Workshop (ORE 2013)*, 2013, p. 94.

[343] M. d’Aquin, J. Davies, and E. Motta, “Smart Cities’ Data: Challenges and Opportunities for Semantic Technologies,” *IEEE Internet Comput.*, no. 6, pp. 66–70, 2015.

[344] Ontotext, “The Truth about Triple Stores.” Ontotext, 2014.

[345] P. Bednar, M. Sarnovsky, and V. Demko, “RDF vs. NoSQL databases for the semantic web applications,” 2014, pp. 361–364.

[346] V. Kilintzis, N. Beredimas, and I. Chouvarda, “Evaluation of the performance of open-source RDBMS and triplestores for storing medical data over a web service,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 4499–4502.

[347] V. Mironov, N. Seethappan, W. Blondé, E. Antezana, B. Lindi, and M. Kuiper, “Benchmarking triple stores with biological data,” *ArXiv Prepr. ArXiv10121632*, 2010.

[348] M. Voigt, A. Mitschick, and J. Schulz, “Yet Another Triple Store Benchmark? Practical Experiences with Real-World Data.,” in *SDA*, 2012, pp. 85–94.

[349] LIS International Summer School, “Library Linked Data.” LIS International Summer School, 2016.

[350] P. Hitzler and K. Janowicz, “Linked Data, Big Data, and the 4th Paradigm.,” *Semantic Web*, vol. 4, no. 3, pp. 233–235, 2013.

[351] C. Bizer, Tom Heath, and Tim Berners-Lee, “Linked Data - The Story So Far,” *Int. J. Semantic Web Inf. Syst.*, 2009.

[352] S. Consoli *et al.*, “Producing Linked Data for Smart Cities: the case of Catania,” *IOS Press*, 2014.

[353] K. Feeney, G. Mendel-Gleason, and R. Brennan, “Linked data schemata: fixing unsound foundations,” *Semantic Web J.-Spec. Issue Qual. Manag. Semantic Web Assets*, 2015.

[354] C. K. Janowicz *et al.*, “Ontology Summit 2014 Communiqué Semantic Web and Big Data Meets Applied Ontology,” 2014.

[355] M. Ali, K. Sriram Bharat, and C. Ranichandra, “Processing RDF Using Hadoop,” in *Advances in Computing and Information Technology: Proceedings of the Second International Conference on Advances in*

*Computing and Information Technology (ACITY) July 13-15, 2012, Chennai, India - Volume 2*, N. Meghanathan, D. Nagamalai, and N. Chaki, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 385–394.

[356] Y. Tian, J. Du, H Wang, Y. Ni, and Y. Yu, “HadoopRDF : A Scalable RDF Data Analysis System,” presented at the International conference on Intelligent Computing Theories and Applications, 2012.

[357] M. F. Husain, P. Doshi, L. Khan, and B. Thuraisingham, “Storage and retrieval of large rdf graph using hadoop and mapreduce,” in *IEEE International Conference on Cloud Computing*, 2009, pp. 680–686.

[358] J. Kawises and W. Vatanawood, “A development of RDF data transfer and query on Hadoop Framework,” in *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on*, 2016, pp. 1–4.

[359] Waves Project, “Waves: Big Data Platform for Real-Time Semantic Stream Management,” Leipzig, 2016.

[360] X. Ren, O. Curé, H. Khrouf, Z. Kazi-Aoul, and Y. Chabchoub, “Apache spark and Apache Kafka at the rescue of distributed RDF stream processing engines,” presented at the CEUR Workshop Proceedings, 2016, vol. 1690.

[361] S. Konstantopoulos, A. Charalambidis, G. Mouchakis, A. Troumpoukis, J. Jakobitch, and V. Karkaletsis, “Semantic Web Technologies and Big Data Infrastructures: SPARQL Federated Querying of Heterogeneous Big Data Stores,” 2016.

[362] D. Ostrowski, N. Rychtyckyj, P. MacNeille, and M. Kim, “Integration of Big Data Using Semantic Web Technologies,” 2016, pp. 382–385.

[363] S. K. Bansal and S. Kagemann, “Integrating Big Data: A Semantic Extract-Transform-Load Framework.,” *IEEE Comput.*, vol. 48, no. 3, pp. 42–50, 2015.

[364] A. Ta’a, M. S. Abdullah, and N. M. Norwawi, “Ontology-Based Extraction-Transformation-Loading (ETL) Processes Model in Data Warehouse Environments,” 2008.

[365] ISO/IEC JTC1, “Smart Cities Preliminary Report,” 2014.

[366] F. T. Fonseca, M. J. Egenhofer, C. A. Davis, and K. A. Borges, “Ontologies and knowledge sharing in urban GIS,” *Comput. Environ. Urban Syst.*, vol. 24, no. 3, pp. 251–272, 2000.

[367] Rosario Uceda-Sosa, Biplav Srivastava, and Robert J. Schloss, “Building a Highly Consumable Semantic Model for Smarter Cities,” 2011.

[368] P. Bellini, M. Benigni, R. Billero, P. Nesi, and N. Rauch, “Km4City ontology building vs data harvesting and cleaning for smart-city services,” *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 827–839, Dec. 2014.

[369] G. Nemirovski, A. Nolle, Á. Sicilia, I. Ballarini, and V. Corado, “Data integration driven ontology design, case study smart city,” in *Proceedings of*

*the 3rd International Conference on Web Intelligence, Mining and Semantics*, 2013, p. 43.

[370] G. Gröger and L. Plümer, “CityGML – Interoperable semantic 3D city models,” *ISPRS J. Photogramm. Remote Sens.*, vol. 71, pp. 12–33, Jul. 2012.

[371] M. Fox, “ISO 37120 indicator URIs.” ISO, 2015.

[372] S. Consoli *et al.*, “A Smart City Data Model based on Semantics Best Practice and Principles,” 2015, pp. 1395–1400.

[373] E. D’Avanzo, A. Lieto, and T. Kuflik, “Manually vs semiautomatic domain specific ontology building,” 2008.

[374] M. Gruninger and M. S. Fox, “Methodology for the Design and Evaluation of Ontologies,” presented at the Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.

[375] G. Schreiber, B. Wielinga, W. Jansweijer, and others, “The KACTUS view on the ‘O’word,” in *IJCAI workshop on basic ontological issues in knowledge sharing*, 1995, pp. 159–168.

[376] M. Fernández-López, “Overview of methodologies for building ontologies,” 1999.

[377] J. R. Punuru, “Machine Learning Techniques for Automatic Ontology Extraction from Domain Texts,” 2007.

[378] K. Liu, W. R. Hogan, and R. S. Crowley, “Natural Language Processing methods and systems for biomedical ontology learning,” *J. Biomed. Inform.*, vol. 44, no. 1, pp. 163–179, Feb. 2011.

[379] S. Chen and M-A. Williams, “Learning Personalized Ontologies from Text: A Review on an Inherently,” *Pers. Inf. Retr. Access Concepts Methods*, p. 1, 2008.

[380] J. Punuru and P. Chen, “Extraction of non-hierarchical relations from domain texts,” in *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, 2007, pp. 444–449.

[381] J. Punuru and J. Chen, “Learning taxonomical relations from domain texts using WordNet and word sense disambiguation,” in *Granular Computing (GrC), 2012 IEEE International Conference on*, 2012, pp. 382–387.

[382] K. Doing-Harris, Y. Livnat, and S. Meystre, “Automated concept and relationship extraction for the semi-automated ontology management (SEAM) system,” *J. Biomed. Semant.*, vol. 6, no. 1, Dec. 2015.

[383] T. Strasser *et al.*, “A Review of Architectures and Concepts for Intelligence in Future Electric Energy Systems,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2424–2438, Apr. 2015.

[384] F. Li *et al.*, “Smart Transmission Grid: Vision and Framework,” *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 168–177, Sep. 2010.

[385] K. Moslehi and R. Kumar, “A Reliability Perspective of the Smart Grid,” *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 57–64, Jun. 2010.

[386] I. McNeil, *An Encyclopedia of the History of Technology*. London: Routledge, 1990.

[387] Museum of scienve and industry, “Timeline,” 2008. [Online]. Available: http://www.mosi.org.uk/collections/explore-the-collections/ferranti-online/timeline.aspx. [Accessed: 14-Aug-2015].

[388] United States Congress, *Public Utility Regulatory Policies Act*. 1978.

[389] G. Simmonds, *Regulation of the UK electricity industry*. University of Bath School of Management, 2002.

[390] “Existing Electric Generating units in the United States.” 2005.

[391] J. Martin, “Distributed vs. centralized electricity generation: are we witnessing a change of paradigm,” *Introd. Distrib. Gener. Paris HEC Www Vernimmen NetftpAnintroductiontodistributedgeneration Pdf*, 2009.

[392] B. F. Hobbs, “Optimization methods for electric utility resource planning,” *Regul. Reg. Power Syst.*, p. 159, 1995.

[393] J. P. Painuly, “Barriers to renewable energy penetration; a framework for analysis,” *Renew. Energy*, vol. 24, no. 1, pp. 73–89, 2001.

[394] REN21, “10 years of renewable energy progress.” 2014.

[395] UNEP, “Agenda 21 - United Nations Environment Programme (UNEP),” 1992. [Online]. Available: http://www.unep.org/Documents.Multilingual/Default.asp?documentid=52. [Accessed: 26-Aug-2015].

[396] U. Nations, “Kyoto Protocol to the United Nations Framework Convention on Climate Change,” 1998.

[397] PetroStrategies, Inc., “Are We Running Out of Oil and Gas?,” 2015. [Online]. Available: http://www.petrostrategies.org/Learning_Center/are_we_running_out_of_oil_and_gas.htm. [Accessed: 13-Aug-2015].

[398] “History of Wind Energy | Department of Energy,” 2014. [Online]. Available: http://energy.gov/eere/wind/history-wind-energy. [Accessed: 14-Aug-2015].

[399] “History of Solar Energy in California - Go Solar California,” 2010. [Online]. Available: http://www.gosolarcalifornia.ca.gov/about/gosolar/california.php. [Accessed: 14-Aug-2015].

[400] M. P. F. Hommelberg, C. J. Warmer, I. G. Kamphuis, J. K. Kok, and G. J. Schaeffer, “Distributed Control Concepts using Multi-Agent technology and Automatic Markets: An indispensable feature of smart power grids,” 2007, pp. 1–7.

[401] M. Manfren, P. Caputo, and G. Costa, “Paradigm shift in urban energy systems through distributed generation: Methods and models,” *Appl. Energy*, vol. 88, no. 4, pp. 1032–1048, Apr. 2011.

[402] A. Y. Saber and G. K. Venayagamoorthy, “Efficient Utilization of Renewable Energy Sources by Gridable Vehicles in Cyber-Physical Energy Systems,” *IEEE Syst. J.*, vol. 4, no. 3, pp. 285–294, Sep. 2010.

[403] Y Kojima, “Operation results of the hachinohe microgrid project,” 5th Microgrid Symposium, 2014.

[404] S. Bae and A. Kwasinski, “Dynamic Modeling and Operation Strategy for a Microgrid With Wind and Photovoltaic Resources,” *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1867–1876, Dec. 2012.

[405] M. E. Torres-Hernández and M. Vélez-Reyes, “Hierarchical control of hybrid power systems,” in *Power Electronics Congress, 2008. CIEP 2008. 11th IEEE International*, 2008, pp. 169–176.

[406] R. H. Lasseter, “MicroGrids,” 2002

[407] P. Stluka, D. Godbole, and T. Samad, “Energy management for buildings and microgrids,” in *Proceedings of the IEEE Conference on Decision and Control*, 2011, pp. 5150–5157.

[408] J. Chen, X. Yang, L. Zhu, and M. Zhang, “Microgrid Multi-objective Economic Operation Optimization Considering Reactive Power,” in *Unifying Electrical Engineering and Electronics Engineering*, vol. 238, 2014, pp. 507–516.

[409] M. Peik-Herfeh, H. Seifi, and M. K. Sheikh-El-Eslami, “Decision making of a virtual power plant under uncertainties for bidding in a day-ahead market using point estimate method,” *Int. J. Electr. Power Energy Syst.*, vol. 44, no. 1, pp. 88–98, Jan. 2013.

[410] F. Pilo, G. Pisano, and G. G. Soma, “Optimal Coordination of Energy Resources With a Two-Stage Online Active Management,” *Ind. Electron. IEEE Trans. On*, vol. 58, no. 10, pp. 4526–4537, Oct. 2011.

[411] C. Abbey, K. Strunz, and G. Joos, “A Knowledge-Based Approach for Control of Two-Level Energy Storage for Wind Energy Systems,” *IEEE Trans. Energy Convers.*, vol. 24, no. 2, pp. 539–547, Jun. 2009.

[412] T. Logenthiran, D. Srinivasan, and D. Wong, “Multi-agent coordination for DER in MicroGrid,” in *Sustainable Energy Technologies, 2008. ICSET 2008. IEEE International Conference on*, 2008, pp. 77–82.

[413] J. Jimeno, J. Anduaga, J. Oyarzabal, and A. G. De Muro, “Architecture of a microgrid energy management system,” *Eur. Trans. Electr. Power*, vol. 21, no. 2, pp. 1142–1158, 2011.

[414] A. L. Dimeas and N. D. Hatziargyriou, “Operation of a Multiagent System for Microgrid Control,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1447–1455, Aug. 2005.

[415] M.-T. Kuo and S.-D. Lu, “Design and Implementation of Real-Time Intelligent Control and Structure Based on Multi-Agent Systems in Microgrids,” *Energies*, vol. 6, no. 11, pp. 6045–6059, Nov. 2013.

[416] R. S. Adhikari, N. Aste, and M. Manfren, “Optimization concepts in district energy design and management – A case study,” *Energy Procedia*, vol. 14, pp. 1386–1391, Jan. 2012.

[417] M. Houwing and M. Ilic, “The value of IT for virtual power plants with micro cogeneration systems,” in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, 2008, pp. 1–6.

[418] K. H. van Dam, M. Houwing, Z. Lukszo, and I. Bouwmans, “Agent-based control of distributed electricity generation with micro combined heat and power - Cross-sectoral learning for process and infrastructure engineers,” *Comput. Chem. Eng.*, vol. 32, no. 1–2, pp. 205–217, 2008.

[419] B. Roossien, M. Hommelberg, C. Warmer, K. Kok, and J.-W. Turkstra, “Virtual power plant field experiment using 10 micro-CHP units at consumer premises,” in *SmartGrids for Distribution, 2008. IET-CIRED. CIRED Seminar*, 2008, pp. 1–4.

[420] M. H. Ali, Bin Wu, and R. A. Dougal, “An Overview of SMES Applications in Power and Energy Systems,” *IEEE Trans. Sustain. Energy*, vol. 1, no. 1, pp. 38–47, Apr. 2010.

[421] I. González, A. Ramiro, M. Calderón, A. J. Calderón, and J. F. González, “Estimation of the state-of-charge of gel lead-acid batteries and application to the control of a stand-alone wind-solar test-bed with hydrogen support,” *Int. J. Hydrog. Energy*, vol. 37, no. 15, pp. 11090–11103, Aug. 2012.

[422] C. Wang and M. H. Nehrir, “Power Management of a Stand-Alone Wind/Photovoltaic/Fuel Cell Energy System,” *IEEE Trans. Energy Convers.*, vol. 23, no. 3, pp. 957–967, Sep. 2008.

[423] A. Pini Prato, F. Strobino, M. Broccardo, and L. Parodi Giusino, “Integrated management of cogeneration plants and district heating networks,” *Appl. Energy*, vol. 97, pp. 590–600, Sep. 2012.

[424] J. Aghaei and M.-I. Alizadeh, “Multi-objective self-scheduling of CHP (combined heat and power)-based microgrids considering demand response programs and ESSs (energy storage systems),” *Energy*, vol. 55, pp. 1044–1054, Jun. 2013.

[425] Mayor of London, “District Heating Manual for London.” Greater London Authority, 2013.

[426] D. Buoro, P. Pinamonti, and M. Reini, “Optimization of a Distributed Cogeneration System with solar district heating,” *Appl. Energy*, vol. 124, pp. 298–308, 2014.

[427] M. Geidl and G. Andersson, “Optimal Coupling of Energy Infrastructures,” in *Power Tech, 2007 IEEE Lausanne*, 2007, pp. 1398–1403.

[428] M. Geidl and G. Andersson, "Optimal Power Flow of Multiple Energy Carriers," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 145–155, Feb. 2007.

[429] A. Sheikhi, A. M. Ranjbar, and H. Oraee, "Financial analysis and optimal size and operation for a multicarrier energy system," *Energy Build.*, vol. 48, pp. 71–78, May 2012.

[430] P. Davidsson and F. Wernstedt, "A multi-agent system architecture for coordination of just-in-time production and distribution," *Knowl. Eng. Rev.*, vol. 17, no. 4, pp. 317–329, Dec. 2002.

[431] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.

[432] M. Parvania and M. Fotuhi-Firuzabad, "Demand Response Scheduling by Stochastic SCUC," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 89–98, Jun. 2010.

[433] A.-H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, and R. Schober, "Optimal and autonomous incentive-based energy consumption scheduling algorithm for smart grid," in *Innovative Smart Grid Technologies (ISGT), 2010*, 2010, pp. 1–6.

[434] A. R. Hota, M. Juvvanapudi, and P. Bajpai, "Issues and solution approaches in PHEV integration to smart grid," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 217–229, Feb. 2014.

[435] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," *Renew. Sustain. Energy Rev.*, vol. 34, pp. 501–516, Jun. 2014.

[436] A. M. A. Haidar, K. M. Muttaqi, and D. Sutanto, "Technical challenges for electric power industries due to grid-integrated electric vehicles in low voltage distributions: A review," *Energy Convers. Manag.*, vol. 86, pp. 689–700, Oct. 2014.

[437] X. Cheng *et al.*, "Electrified Vehicles and the Smart Grid: The ITS Perspective," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1388–1404, Aug. 2014.

[438] K. Zhang *et al.*, "Optimal decentralized valley-filling charging strategy for electric vehicles," *Energy Convers. Manag.*, vol. 78, pp. 537–550, Feb. 2014.

[439] H. Morais, T. Sousa, Z. Vale, and P. Faria, "Evaluation of the electric vehicle impact in the power demand curve in a smart grid environment," *Energy Convers. Manag.*, vol. 82, pp. 268–282, Jun. 2014.

[440] A. Di Giorgio, F. Liberati, and S. Canale, "Electric vehicles charging control in a smart grid: A model predictive control approach," *Control Eng. Pract.*, vol. 22, pp. 147–162, Jan. 2014.

[441] A. H. Kazmi, M. J. O'grady, D. T. Delaney, A. G. Ruzzelli, and G. M. O'hare, "A review of wireless-sensor-network-enabled building energy management systems," *ACM Trans. Sens. Netw. TOSN*, vol. 10, no. 4, p. 66, 2014.

[442] A. De Paola, M. Ortolani, G. Lo Re, G. Anastasi, and S. K. Das, “Intelligent Management Systems for Energy Efficiency in Buildings: A Survey,” *ACM Comput. Surv.*, vol. 47, no. 1, pp. 1–38, Jun. 2014.

[443] J. A. Clarke *et al.*, “Simulation-assisted control in building energy management systems,” *Energy Build.*, vol. 34, no. 9, pp. 933–940, 2002.

[444] A. Costa, R. S. Garay, T. Messervey, and M. M. Keane, “Value of building simulation in sport facilities operation,” in *12th Conference of International Building Performance Simulation Association*, 2011.

[445] Z. Qian, C. C. Seepersad, V. R. Joseph, J. K. Allen, and C. F. J. Wu, “Building Surrogate Models Based on Detailed and Approximate Simulations,” *J. Mech. Des.*, vol. 128, no. 4, p. 668, 2006.

[446] H. Doukas, K. D. Patlitzianas, K. Iatropoulos, and J. Psarras, “Intelligent building energy management system using rule sets,” *Build. Environ.*, vol. 42, no. 10, pp. 3562–3569, Oct. 2007.

[447] S. Atthajariyakul and T. Leephakpreeda, “Neural computing thermal comfort index for HVAC systems,” *Energy Convers. Manag.*, vol. 46, no. 15–16, pp. 2553–2565, Sep. 2005.

[448] H. Hagras, V. Callaghan, M. Colley, and G. Clarke, “A hierarchical fuzzy–genetic multi-agent architecture for intelligent buildings online learning, adaptation and control,” *Inf. Sci.*, vol. 150, no. 1, pp. 33–57, 2003.

[449] R. Alcalá, J. M. Benítez, J. Casillas, O. Cordón, and R. Pérez, “Fuzzy control of HVAC systems optimized by genetic algorithms,” *Appl. Intell.*, vol. 18, no. 2, pp. 155–177, 2003.

[450] Y. Yuan, J. Yuan, H. Du, and L. Li, “An improved multi–objective ant colony algorithm for building life cycle energy consumption optimisation,” *Int. J. Comput. Appl. Technol.*, vol. 43, no. 1, pp. 60–66, 2012.

[451] K. Li, H. Su, and J. Chu, “Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study,” *Energy Build.*, vol. 43, no. 10, pp. 2893–2899, Oct. 2011.

[452] S. Mitra and Y. Hayashi, “Neuro-fuzzy rule generation: survey in soft computing framework,” *IEEE Trans. Neural Netw.*, vol. 11, no. 3, 2000.

[453] D. Kolokotsa, G. S. Stavrakakis, K. Kalaitzakis, and D. Agoris, “Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks,” *Eng. Appl. Artif. Intell.*, vol. 15, no. 5, pp. 417–428, Sep. 2002.

[454] P. May-Ostendorp, G. P. Henze, C. D. Corbin, B. Rajagopalan, and C. Felsmann, “Model-predictive control of mixed-mode buildings with rule extraction,” *Build. Environ.*, vol. 46, no. 2, pp. 428–437, Feb. 2011.

[455] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, “State of the art in building modelling and energy performances prediction: A review,” *Renew. Sustain. Energy Rev.*, vol. 23, pp. 272–288, Jul. 2013.

[456] Mona Visa, “Mona Visa Dashboard,” 2016. [Online]. Available: http://www.monavisa.info/home/dashboard.jsp. [Accessed: 30-Dec-2015].

[457] Plugwise, “The Plugwise app,” 2016. [Online]. Available: https://www.plugwise.com/plugwise-app. [Accessed: 16-Nov-2016].

[458] J. Keirstead, M. Jennings, and A. Sivakumar, “A review of urban energy system models: Approaches, challenges and opportunities,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3847–3866, Aug. 2012.

[459] J. Lagorse, D. Paire, and A. Miraoui, “A multi-agent system for energy management of distributed power sources,” *Renew. Energy*, vol. 35, no. 1, pp. 174–182, Jan. 2010.

[460] W. Radziszewska, Z. Nahorski, M. Parol, and P. Palka, “Intelligent Computations in an Agent-Based Prosumer-Type Electric Microgrid Control System,” in *Issues and Challenges of Intelligent Systems and Computational Intelligence*, 2014.

[461] J. K. Kok, C. J. Warmer, and I. G. Kamphuis, “PowerMatcher: Multi-Agent Control in the Electricity Infrastructure,” in *Autonomous Agents and Multi-Agent Systems*, 2005.

[462] J. Kok, “The PowerMatcher: Smart Coordination for the Smart Electricity Grid,” 2013.

[463] M. P. F. Hommelberg, B. J. van der Velde, C. J. Warmer, I. G. Kamphuis, and J. K. Kok, “A novel architecture for real-time operation of multi-agent based coordination of demand and supply,” in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008, pp. 1–5.

[464] T. Logenthiran, D. Srinivasan, and A. M. Khambadkone, “Multi-Agent System for Energy Resource Scheduling of Integrated Microgrids in a Distributed System,” *Electr. Power Syst. Res.*, vol. 81, pp. 138–148, 2011.

[465] O. van Pruissen, V. Kamphuis, A. van der Togt, and E. Werkman, “A Thermal grid coordinated by a multi agent energy management system,” in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, 2013, pp. 1–5.

[466] K. Kok *et al.*, “Intelligent energy management using power matcher: recent results from field deployments and simulation studies,” 2013.

[467] C.-X. Dou and B. Liu, “Hierarchical management and control based on MAS for distribution grid via intelligent mode switching,” *Int. J. Electr. Power Energy Syst.*, vol. 54, pp. 352–366, 2014.

[468] K. H. van Dam, “Capturing socio-technical systems with agent-based modelling,” Delft University of Technology, Delft, the Netherlands, 2009.

[469] A. W. McMorran, “An Introduction to IEC 61970-301 and 61968-11: The Common Information Model,” University of Strathclyde, 2007.

[470] OpenADR Alliance, “OpenADR 2.0 Profile Specification A Profile,” 20110712–1, 2011.

[471] Energy@home Association, “Energy@home data Model,” 2015.

[472] P. Pauwels and W. Terkaj, “EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology,” *Autom. Constr.*, vol. 63, pp. 100–133, Mar. 2016.

[473] L. Daniele, F. den Hartog, and J. Roes, “Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology,” in *Formal Ontologies Meet Industry*, vol. 225, R. Cuel and R. Young, Eds. Cham: Springer International Publishing, 2015, pp. 100–112.

[474] Bristish Standards, “Energy management system application program interface (EMS-API) - Part 301: Common information model (CIM) base (IEC 61970-301:2011),” British Standards Institution, Nov. 2011.

[475] International Electrotechnical COmmission, “Energy management system application program interface (EMS-API) - Part 501: Common Information Model Resource Description Framework (CIM RDF) schema.” 2006.

[476] C. Wang, H. Liu, and F. Wu, “The extend CIM for MicroGrid,” in *Electricity Distribution (CICED), 2012 China International Conference on*, 2012, pp. 1–5.

[477] T. H. Beach, Y. Rezgui, H. Li, and T. Kasim, “A rule-based semantic approach for automated regulatory compliance in the construction sector,” *Expert Syst. Appl.*, vol. 42, no. 12, pp. 5219–5231, Jul. 2015.

[478] M. Kadolsky, P. Katranuschkov, M. Dolenc, K. Baumgärtel, R. Klinc, and G. Gudnason, *ISES Deliverable D3. 1: Ontology specification*. 2014.

[479] J. Ploennigs, H. Diblowski, A. Roder, K. Kabitzsch, and B. Hensel, “Deliverable D4.1: ICT PLATFORM FOR HOLISTIC ENERGY EFFICIENCY SIMULATION AND LIFECYCLE MANAGEMENT OF PUBLIC USE FACILITIES.” 2011.

[480] RESILIENT, “D2.2 - District Information Model.pdf.” 2013.

[481] C. Derksen, C. Branki, and R. Unland, “A framework for agent-based simulations of hybrid energy infrastructures,” in *2012 Federated Conference on Computer Science and Information Systems, FedCSIS 2012*, 2012, pp. 1293–1299.

[482] Jean-Marie Bahu and Romain Nouvel, “Development of the CItyGML ADE Energy,” 2015.

[483] J. Keirstead, N. Samsatli, and Prof. Nilay Shah, “SynCity: An inegrated tool kit for urban enegy systems modelling,” presented at the Urban Research Symposium, 2009.

[484] S. D. McArthur, S. M. Strachan, and G. Jahn, “The design of a multi-agent transformer condition monitoring system,” *Power Syst. IEEE Trans. On*, vol. 19, no. 4, pp. 1845–1852, 2004.

[485] J. A. Hossack, J. Menal, S. D. J. McArthur, and J. R. McDonald, “A multiagent architecture for protection engineering diagnostic assistance,” *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 639–647, May 2003.

[486] J. Keirstead and K. H. van Dam, “A comparison of two ontologies for agent-based modelling of energy systems,” in *Proceedings of the First International Workshop on Agent Technologies for Energy Systems (ATES 2010)*, 2010, pp. 21–28.

[487] V. M. Catterson, E. M. Davidson, and S. D. J. McArthur, “Issues in integrating existing multi-agent systems for power engineering applications,” in *Intelligent Systems Application to Power Systems, 2005. Proceedings of the 13th International Conference on*, 2005, p. 6 pp.

[488] OGC, “CityGML UtilityNetworkADE - Citygml,” 2012. [Online]. Available: http://www.citygmlwiki.org/index.php/CityGML_UtilityNetworkADE. [Accessed: 18-Jan-2016].

[489] SWAN, “What are smart water networks.” 2010.

[490] M. Allen, A. Preis, M. Iqbal, and A. J. Whittle, “Water Distribution System Monitoring and Decision Support Using a Wireless Sensor Network,” 2013, pp. 641–646.

[491] F. Halters, E. Zondervan, and A. de Haan, “Integrated optimization of a waste water treatment plant using statistical analysis,” *J. Hazard. Mater.*, vol. 179, no. 1–3, pp. 480–487, Jul. 2010.

[492] C. Schenk, “A systems-Based Generic Decision Support System Application to Urban Water Management,” ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2010.

[493] CTRL+SWAN, “Cloud Technologies & ReaL time monitoring + Smart WAter Network,” 2016. [Online]. Available: http://www.eip-water.eu/CTRL_SWAN. [Accessed: 18-Feb-2016].

[494] EIP Water, “AugMent - Water Monitoring for Decision Support (AG124) | EIP Water,” 2016. [Online]. Available: http://www.eip-water.eu/AugMent. [Accessed: 18-Feb-2016].

[495] A. Cahn, “Shaping the architecture of smart water networks.” Water & Sewerage Journal, 2013.

[496] M. S. Babel, M. S. Islam, and A. Das Gupta, “Leakage management in a low-pressure water distribution network of Bangkok,” *Water Sci. Technol. Water Supply*, vol. 9, no. 2, p. 141, Jun. 2009.

[497] S. R. Mounce, C. Pedraza, T. Jackson, P. Linford, and J. B. Boxall, “Cloud Based Machine Learning Approaches for Leakage Assessment and Management in Smart Water Networks,” *Procedia Eng.*, vol. 119, pp. 43–52, 2015.

[498] S. W. Lee, S. Sarp, D. J. Jeon, and J. H. Kim, “Smart water grid: the future water management platform,” *Desalination Water Treat.*, vol. 55, no. 2, pp. 339–346, Jul. 2015.

[499] D. Kenny, “Smart Water Network Monitoring SBWWI - Intelligent Networks,” Future Water Association, 2013.

[500] J. M Miller and Mark Leinmiller, “Why Smart Water Networks Boost Efficiency,” Schneider Electric, 2014.

[501] F Williamson *et al.*, “Online water quality monitoring in the distribution network,” *Water Pract. Technol.*, vol. 9, no. 4, 2014.

[502] T. Robles *et al.*, “An IoT based reference architecture for smart water management processes,” *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl. JoWUA*, vol. 6, no. 1, pp. 4–23, 2015.

[503] A. Hauser, “Communication in Smart Water Networks, Workgroup Interoperability,” presented at the SWAN, 2016.

[504] CUAHSI, “Hydrologic Information System - Hydrologic Ontology for Discovery,” 2008. [Online]. Available: http://his.cuahsi.org/ontologyfiles.html. [Accessed: 23-Oct-2015].

[505] R. Raskin, “Semantic Web for Earth and Environmental Terminology,” NASA, 2011.

[506] L.M Vilches-Blázquez, J. A. Ramos, F. J. López-Pellicer, O. Corcho, and J. Nogueras-Iso, “hydrOntology: a global ontology of the hydrographical domain,” 2009. [Online]. Available: http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/ontologies/107-hydrontology. [Accessed: 23-Oct-2015].

[507] N. Fleischmann, *EIP Thesaurus for Matchmaking*. EIP Water, 2013.

[508] OGC, “WaterML2 - A Global Standard for Hydrological Time Series,” OGC, 2014.

[509] Australian Bureau of Meteorology, “Water Data Transfer Format,” Australian Bureau of Meteorology, 2013.

[510] WatERP, “D1.1 Generic taxonomy for water supply distribution chain.” 2013.

[511] L. Reynolds, “SWIM – a semantic ontology for interoperability.” 2014.

[512] INSPIRE, “D2.8.III.6 Data Specification on Utility and Government Services – Technical Guidelines,” European Commission Joint Research Centre, Dec. 2013.

[513] Grow-smarter, “Barcelona blog #4: passive refurbishments in the residential sector and semantic data,” Oct-2016. [Online]. Available: http://www.grow-smarter.eu/inform/blog-updates/blog-archive/?c=search&uid=bbe537b4. [Accessed: 29-Nov-2016].

[514] S. Howell, Y. Rezgui, T. Beach, W. Zhao, J. Terlet, and H. Li, “Smart Water System Interoperability: Integrating Data and Analytics for Demand Optimized Management through Semantics,” 2016.

[515] L. Sanchez *et al.*, “SmartSantander: IoT experimentation over a smart city testbed,” *Comput. Netw.*, vol. 61, pp. 217–238, Mar. 2014.

[516] R. Peinl, “Semantic Web: State of the Art and Adoption in Corporations,” *KI - Künstl. Intell.*, vol. 30, no. 2, pp. 131–138, Jun. 2016.

[517] Harbor Research, “Smarter SCADA & The Internet of Things,” Smarter SCADA Workshop, 2012.

[518] M. N. K. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students*, 5th ed. New York: Prentice Hall, 2009.

[519] N. M. Lemos, *An introduction to the theory of knowledge*. Cambridge, UK; New York: Cambridge University Press, 2007.

[520] D. E. Gray, *Doing research in the real world*. Sage, 2013.

[521] N. Rescher, *Epistemology: an introduction to the theory of knowledge*. Albany: State University of New York, 2003.

[522] Paul Moser, *The Oxford Handbook of Epistemology*. Oxford University Press, 2005.

[523] S. Olbrich, B. Mueller, and F. Niederman, “Theory Emergence in IS Research: The Grounded Theory Method Applied,” in *10. JAIS Theory Development Workshop*, 2011.

[524] R. Cameron, “A sequential mixed model research design: Design, analytical and display issues,” *Int. J. Mult. Res. Approaches*, vol. 3, no. 2, pp. 140–152, 2009.

[525] J. Ritchie and J. Lewis, Eds., *Qualitative research practice: a guide for social science students and researchers*. London ; Thousand Oaks, Calif: Sage Publications, 2003.

[526] N. K Dash, “Research Methods Resource - Selection of the Research Paradigm and Methodology,” *IGNOU*, 2005. [Online]. Available: http://www.celt.mmu.ac.uk/researchmethods/Modules/Selection_of_methodology/. [Accessed: 04-Sep-2016].

[527] W. J. Orlikowski and J. J. Baroudi, “Studying information technology in organizations: Research approaches and assumptions,” *Inf. Syst. Res.*, vol. 2, no. 1, pp. 1–28, 1991.

[528] S. Demeyer, “Research Methods in Computer Science,” Universiteit Antwerpen, 2010.

[529] M. Tedre, “Know your discipline: Teaching the philosophy of computer science,” *J. Inf. Technol. Educ.*, vol. 6, no. 1, pp. 105–122, 2007.

[530] G. Goldkuhl, “Meanings of Pragmatism: Ways to conduct information systems research,” *Action Lang. Organ. Inf. Syst.*, 2004.

[531] G. C. Peng, J. M. B. Nunes, and F. Annansingh, “Investigating information systems with mixed-methods research,” in *Proceedings of the IADIS International Workshop on Information Systems Research Trends, Approaches and Methodologies*, 2011.

[532] M. Easterby-Smith, R. Thorpe, and A. Lowe, *Management Research: An Introduction*. SAGE Publications, 2002.

[533] L. K. Soiferman, “Compare and Contrast Inductive and Deductive Research Approaches.,” *Online Submiss.*, 2010.

[534] C. P. Pathirage, R. D. G. Amaratunga, and R. P. Haigh, “The role of philosophical context in the development of research methodology and theory,” *Built Hum. Environ. Rev.*, vol. 1, no. 1, pp. 1–10, 2008.

[535] G. Maccani, B. Donnellan, and M. Helfert, “SYSTEMATIC PROBLEM FORMULATION IN ACTION DESIGN RESEARCH: THE CASE OF SMART CITIES,” presented at the ECIS, 2014.

[536] K. Herr and G. L. Anderson, *The action research dissertation: A guide for students and faculty*. Sage publications, 2014.

[537] P. Marshall, K. de Salas, and J. McKay, “Action research in practice: balancing the dual imperatives,” *ACIS 2006 Proc.*, p. 10, 2006.

[538] R. L. Baskerville, “Investigating information systems with action research,” *Commun. AIS*, vol. 2, no. 3es, p. 4, 1999.

[539] William Foote Whyte, et al., *Participatory action research.pdf*. Sage Publications, 1991.

[540] N. Kock, “The three threats of action research: a discussion of methodological antidotes in the context of an information systems study,” *Decis. Support Syst.*, vol. 37, no. 2, pp. 265–286, May 2004.

[541] J. Iivari, “Two strategies for design science research,” Working Paper, 2012 (https:// www. researchgate. net/ profile/ Juhani_ Iivari/ contributions, 2012.

[542] S. Wang and H. Wang, “Towards innovative design research in information systems,” *J. Comput. Inf. Syst.*, vol. 51, no. 1, pp. 11–18, 2010.

[543] Peter Nielsen, Peter Axel, and John Stouby, “Engaged Problem Formulation in IS Research,” *Commun. Assoc. Inf. Syst.*, 2016.

[544] Donald Cooper and Pamela Schindler, *Business Resaerch Methods*, 10th ed. Boston, MA: McGraw-Hill, 2008.

[545] D. F. Polit and C. T. Beck, “Generalization in quantitative and qualitative research: Myths and strategies,” *Int. J. Nurs. Stud.*, vol. 47, no. 11, pp. 1451–1458, Nov. 2010.

[546] A. Gerber, P. Kotze, and A. Van der Merwe, “Design Science Research as Research Approach in Doctoral Studies,” 2015.

[547] J-L. H, Y. Rezgui, H. Li, B. Jayan, and S. Howell, “Ontology-driven development of web services to support district energy applications,” *Unpubl. Manuscr.*, 2017.

[548] ICAM, “ICAM Architecture Part II-Volume IV - Function Modeling Manual (IDEF0),” 1982.

[549] ISO, "ISO/IEC 19501:2005." 2005.

[550] K. H. van Dam, *Capturing socio-technical systems with agent-based modelling.* Delft: Next Generation Infrastructures Foundation, 2009.

[551] Adam Pease, "The Suggested Upper Merged Ontology (SUMO) - Ontology Portal," 2014. [Online]. Available: http://www.adampease.org/OP/. [Accessed: 27-Oct-2015].

[552] USEPA, "EPANET | Water Research | US EPA," 2000. [Online]. Available: https://www.epa.gov/water-research/epanet. [Accessed: 19-May-2016].

[553] RDFlib, "GitHub - RDFLib/rdflib: RDFLib is a Python library for working with RDF, a simple yet powerful language for representing information.," 2008. [Online]. Available: https://github.com/RDFLib/rdflib. [Accessed: 19-May-2016].

[554] Apache Jena, "Jena Ontology API," 2011. [Online]. Available: https://jena.apache.org/documentation/ontology/. [Accessed: 24-Feb-2015].

[555] "protégé," 1999. [Online]. Available: http://protege.stanford.edu/. [Accessed: 01-Jun-2016].

[556] Yvonna Lincoln and Egon Guba, *Naturalistic Inquiry*. Newbury Park, CA: SAGE Publications, 1985.

[557] P. Cronin, F. Ryan, and M. Coughlan, "Undertaking a literature review: a step-by-step approach," *Br. J. Nurs.*, vol. 17, no. 1, pp. 38–43, Jan. 2008.

[558] P. A. Williams, "Making Research Real: Is Action Research a Suitable Methodology for Medical Information Security Investigations?," in *Australian Information Security Management Conference*, 2006, p. 82.

[559] D. Champion and F. A. Stowell, "Validating action research field studies: PEArL," *Syst. Pract. Action Res.*, vol. 16, no. 1, pp. 21–36, 2003.

[560] J. Á. Carvalho, "Validation criteria for the outcomes of design research," in *Pre-ECIS workshop on IT Artefact Design & Workpractice Intervention*, 2012.

[561] C. C. Seepersad, K. Pedersen, J. Emblemsv\a ag, R. Bailey, J. K. Allen, and F. Mistree, *The validation square: how does one verify and validate a design method?* ASME Press, New York, 2006.

[562] G. Dodig-Crnkovic, "Professional ethics in computing and intelligent systems," in *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence*, 2006, pp. 25–27.

[563] M. A. Pierce and J. W. Henry, "Computer ethics: The role of personal, informal, and formal codes," *J. Bus. Ethics*, vol. 15, no. 4, pp. 425–437, 1996.

[564] Cardiff University, "Research Integrity & Governance Code of Practice," 2015.

[565] H. Wicaksono, S. Rogalski, and E. Kusnady, “Knowledge-based intelligent energy management using building automation system,” in *2010 Conference Proceedings IPEC*, 2010, pp. 1140–1145.

[566] H. Wicaksono, P. Dobreva, P. Häfner, and S. Rogalski, “Methodology to Develop Ontological Building Information Model for Energy Management System in Building Operational Phase,” in *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, vol. 454, A. Fred, J. L. G. Dietz, K. Liu, and J. Filipe, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 168–181.

[567] B. Yuce and Y. Rezgui, “An ANN-GA Semantic Rule-Based System to Reduce the Gap Between Predicted and Actual Energy Consumption in Buildings,” *IEEE Trans. Autom. Sci. Eng.*, vol. PP, no. 99, pp. 1–13, 2015.

[568] Preslava Krahtova, Kris McGlinn, Hendro Wicaksono, and Polina Hafner, “Occupancy Modelling and Extended Ontology population,” EC FP7 Project, Project deliverable Deliverable 1.2_version 2, 2013.

[569] Smart Energy Collective, “Introduction to the Universal Smart Energy Framework,” Smart Energy Collective, Arnhem, The Netherlands, 2013.

[570] “NeOn Wiki,” 2014. [Online]. Available: http://neon-toolkit.org/wiki/Main_Page. [Accessed: 10-Apr-2015].

[571] Assela Pathirana, “EPANETTOOLS 0.9.2 : Python Package Index,” Nov-2016. [Online]. Available: https://pypi.python.org/pypi/EPANETTOOLS. [Accessed: 22-Mar-2017].

[572] Google Developers, “Google Maps JavaScript API,” no date. [Online]. Available: https://developers.google.com/maps/documentation/javascript/. [Accessed: 11-Sep-2016].

[573] Cesium Consortium, “Cesium - WebGL Virtual Globe and Map Engine,” 2012. [Online]. Available: https://cesiumjs.org/. [Accessed: 17-Mar-2017].

[574] Andy Ward, Claudio Benghi, Martin Cerny, and Steve Lockley, “The xBIM Toolkit,” 2006. [Online]. Available: https://xbim.codeplex.com/. [Accessed: 17-Mar-2017].

[575] Ruben Verborgh, “N3.js: Lightning fast, spec-compatible, streaming RDF for JavaScript,” 2011. [Online]. Available: https://github.com/RubenVerborgh/N3.js/. [Accessed: 17-Mar-2017].

[576] Chart.js open source project, “Chart.js | Open source HTML5 Charts for your website,” 2013. [Online]. Available: http://www.chartjs.org/. [Accessed: 17-Mar-2017].

[577] apstex, “IFC TOOLS PROJECT,” 2013. [Online]. Available: http://www.ifctoolsproject.com/. [Accessed: 17-Mar-2017].

[578] Claus Nagel, “citygml4j · GitHub,” 2008. [Online]. Available: https://github.com/citygml4j. [Accessed: 17-Mar-2017].

[579] W3C, “SPARQL Query Language for RDF,” 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query. [Accessed: 24-Feb-2015].

[580] University of Geneva, “ICLE Ontologies by ISS@CUI,” 2013. [Online]. Available: http://cui.unige.ch/isi/icle-wiki/ontologies. [Accessed: 19-Feb-2017].

[581] A. Gómez-Péreza, “OOPS!(OntOlogy Pitfall Scanner!): supporting ontology evaluation on-line,” *IOS Press*, p. 2009.

[582] R. G. Raskin and M. J. Pan, “Knowledge representation in the semantic web for Earth and environmental terminology (SWEET),” *Comput. Geosci.*, vol. 31, no. 9, pp. 1119–1125, Nov. 2005.

[583] Complexible, “GitHub - Complexible/pellet,” 2013. [Online]. Available: https://github.com/Complexible/pellet. [Accessed: 19-May-2016].

[584] VITAL, “D3.1.1: Virtual models, data and metadata for ICOs.” 2014.

[585] M. d’Aquin, A. Adamou, E. Daga, S. Liu, K. Thomas, and E. Motta, “Dealing with Diversity in a Smart-City Datahub,” presented at the Proceedings of the Fifth Workshop on Semantics for Smarter Cities, Riva del Garda, Italy, 2014.

[586] BuildingSmart, “IFC4 Add1 Release,” 2015. [Online]. Available: http://www.buildingsmart-tech.org/specifications/ifc-releases/ifc4-add1-release. [Accessed: 18-Jan-2016].

[587] eeBUS, “About us,” 2012. [Online]. Available: https://www.eebus.org/en/about-us/. [Accessed: 01-Jun-2016].

[588] Project Haystack, “Home – Project Haystack,” 2011. [Online]. Available: http://project-haystack.org/. [Accessed: 01-Jun-2016].

[589] Oracle, “Oracle Spatial and Graph: Benchmarkign a Trillion Edges RDF Graph,” Oracle, Redwood Shored, CA, USA, White paper, 2016.

[590] AllegroGraph, “AllegroGraph RDFStore Benchmark Results,” 2017. [Online]. Available: http://franz.com/agraph/allegrograph/agraph_benchmarks.lhtml. [Accessed: 09-Mar-2017].

# 8 APPENDICES

## 8.1 APPENDIX A: SMART WATER ONTOLOGY

During the action research iteration within the water domain, a comprehensive water ontology was developed. As well as the description in section 4.3.4, Figure 92 - Figure 95 illustrate this model to greater depth.

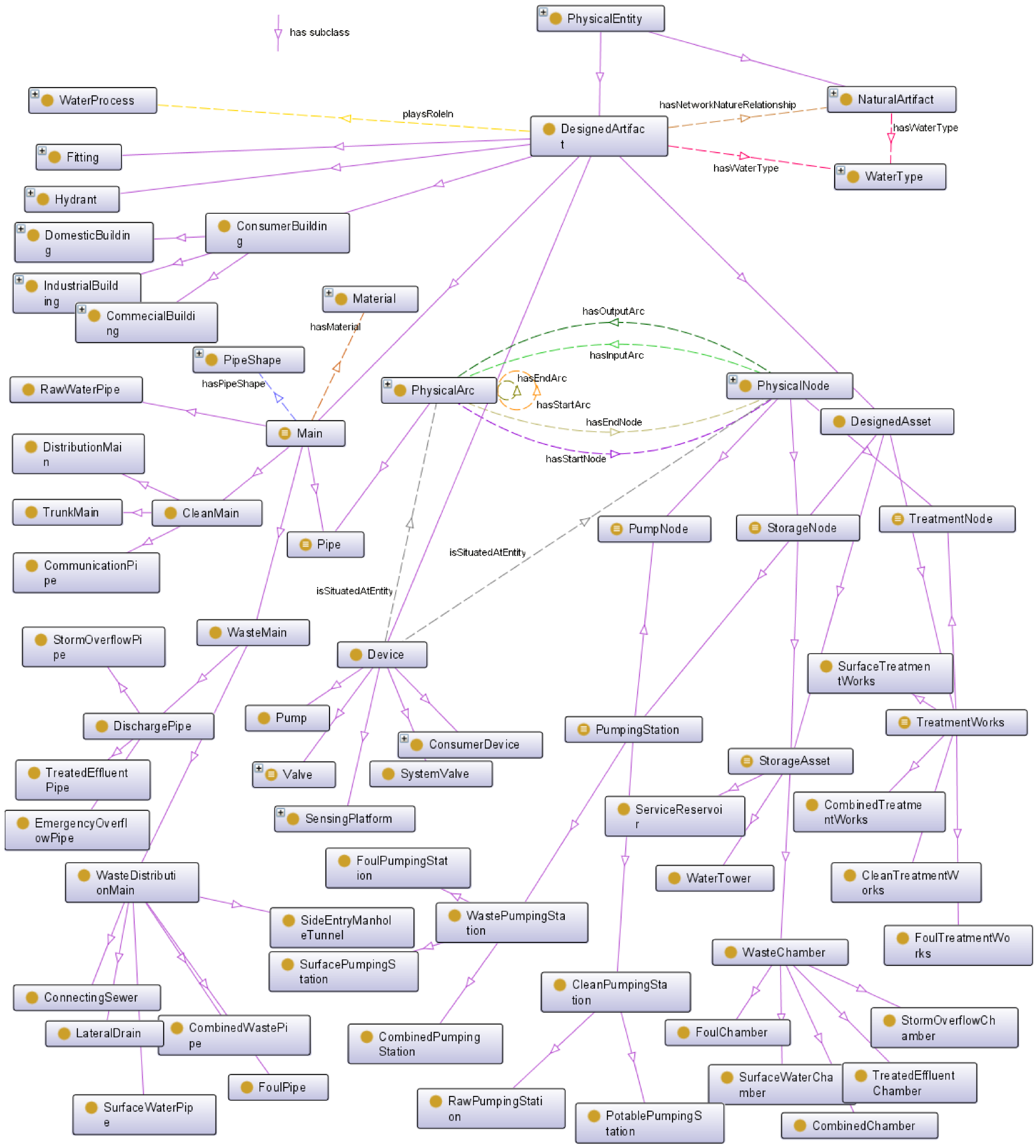


Figure 92: Main supply side WCIM classes and relationships

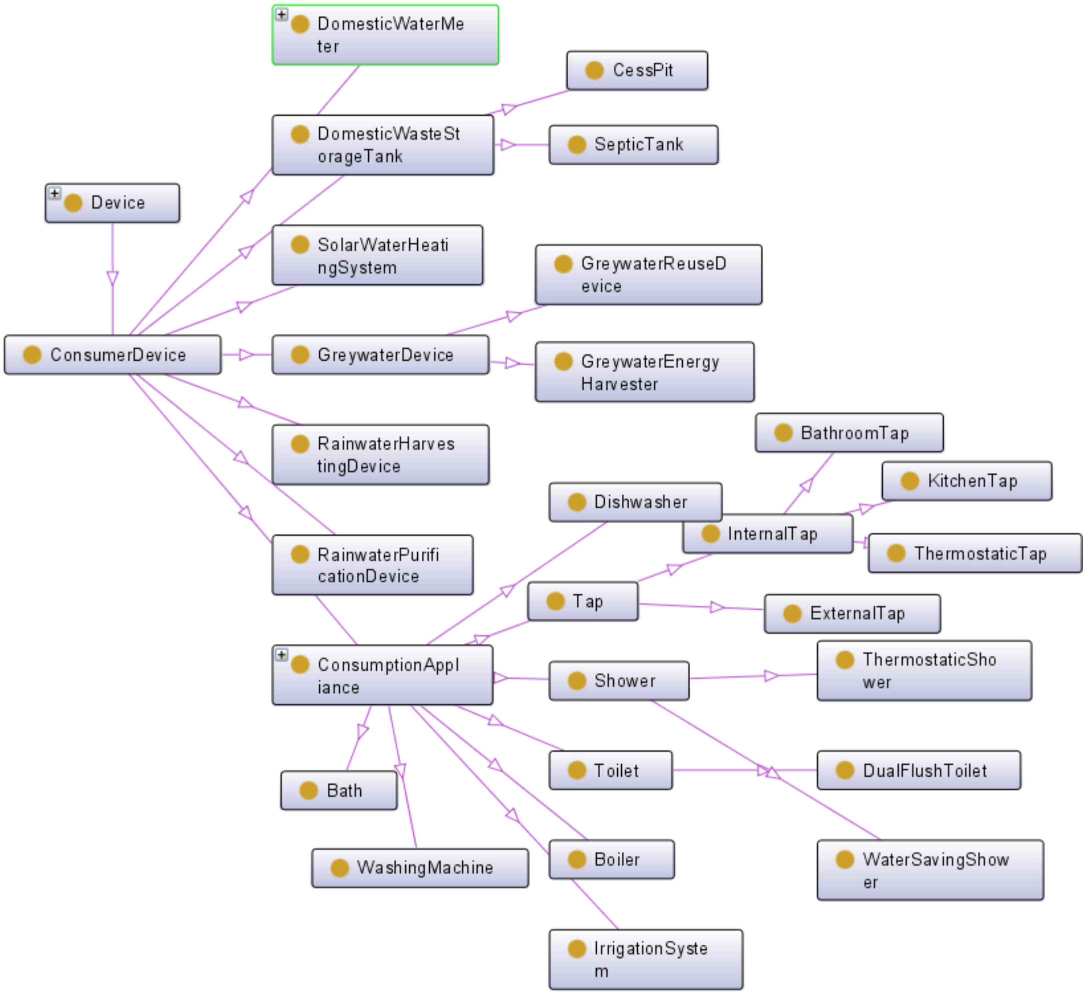


**Figure 93: WCIM Domestic artefacts**

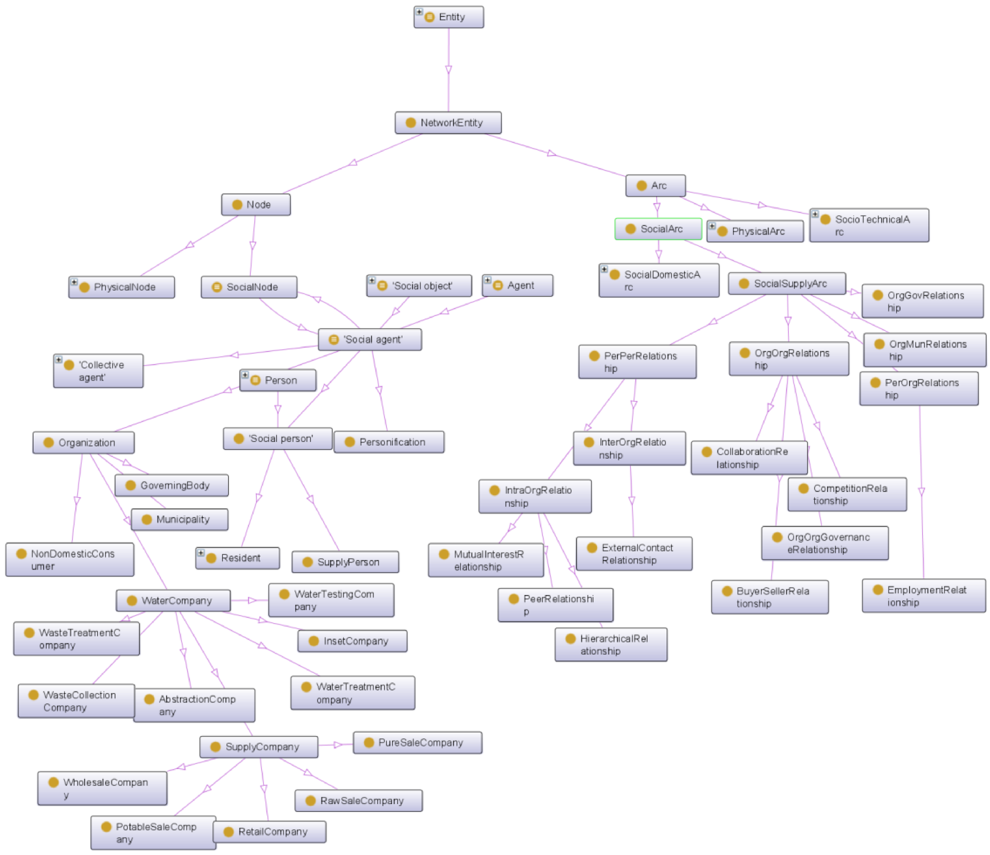


**Figure 94: WVCSM class hierarchy of main social network entities**

| Data Property | Func | Domain | Range |
|---|---|---|---|
| topDataProperty | ☐ | | |
| hasDomicileProperty | ☐ | | |
| hasNumberOfSinksPerDayForDishes | ☑ | Domicile | int |
| hasLawn | ☑ | Domicile | boolean |
| hasNumberOfFlushesPerPersonDay | ☑ | Domicile | int |
| hasAverageMinutesPerShower | ☑ | Domicile | int |
| hasNumberOfShowersPerDay | ☑ | Domicile | int |
| hasNumberOfBathsPerDay | ☑ | Domicile | int |
| hasNumberOfRooms | ☑ | Domicile | int |
| hasResidentProperty | ☐ | | |
| hasAge | ☑ | Resident | int |
| wouldBeCarefulIfAware | ☑ | Resident | boolean |
| willingToPayForDevice | ☑ | Resident | boolean |
| willingToInstallDevice | ☑ | Resident | boolean |
| willingToChange | ☑ | Resident | boolean |
| wantsTipsOnConservation | ☑ | Resident | boolean |
| usesWashingMachineTemp | ☑ | Resident | int |
| usesThermostaticShowerTemp | ☑ | Resident | int |
| reportedWashingMachineLoadsWeekly | ☑ | Resident | int |
| reportedGardenWateringInstancesWeekly | ☑ | Resident | int |
| reportedDishwasherLoadsWeekly | ☑ | Resident | int |
| reportedDaysBetweenSupplierInteraction | ☑ | Resident | int |
| reportedDaysBetweenConsumptionChecks | ☑ | Resident | int |
| reportedAverageBill | ☑ | Resident | double |
| isDoingEnoughToSaveWater | ☑ | Resident | boolean |
| isAwareOfConsumption | ☑ | Resident | boolean |
| hasDOB | ☑ | Resident | date |
| givesWaterConservationImportance | ☑ | Resident | double |
| getsWaterUpdates | ☑ | Resident | boolean |
| feelsEncouragedToSaveWater | ☑ | Resident | boolean |
| couldBeDoingMoreToSaveWater | ☑ | Resident | boolean |
| considersWaterConservationImportant | ☑ | Resident | boolean |
| checksWaterQuality | ☑ | Resident | boolean |
| agreesWithMetering | ☑ | Resident | boolean |

| Object Property | Domain | Range |
|---|---|---|
| hasResidentProperty | | |
| hasWaterSupplier | Resident | SupplyCompany |
| hasConsumptionPattern | Resident | ConsumptionPattern |
| hasBillPayingMethod | Resident | BillPayingMethod |
| wantsUpdatesThroughMethod | Resident | CommunicationMethod |
| hasWaterSavingHabit | Resident | WaterSavingHabit |
| hasHighestEducationLevel | Resident | EducationLevel |
| hasEmploymentStatus | Resident | EmploymentStatus |
| hasGenderIdentity | Resident | GenderIdentity |
| interactsWithSupplierThrough | Resident | CommunicationMethod |
| hasWaterSavingMotivation | Resident | WaterSavingMotivation |

**Figure 95: WVCSM domestic properties**

Figure 96 clarifies the modelling pattern adopted for the WISDOM data layer in accordance with the SSN ontology, by illustrating the mapping between terms used in the ontology and terms used in the KairosDB system.

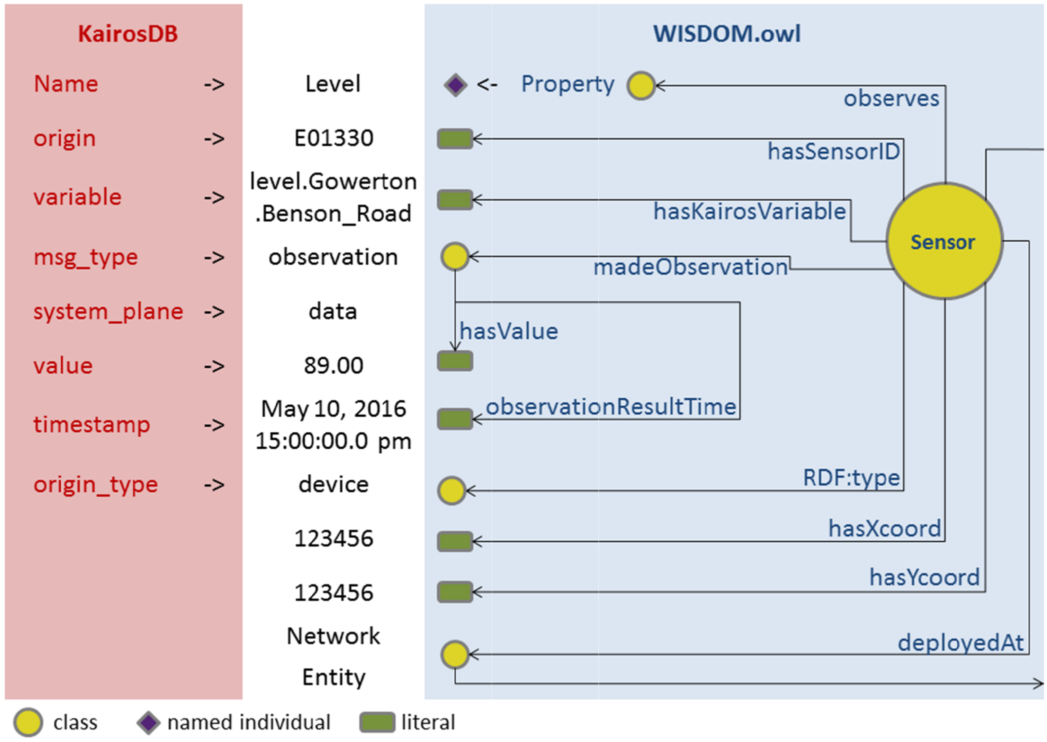


**Figure 96: Mapping between KairosDB terms and WISDOM ontology terms**

The main alignments and points of heterogeneity of the WISDOM ontology with other models are presented in Table 49 - Table 52.

**Table 50: Alignment of WISDOM and WatERP concepts**

| WISDOM concepts | WatERP concept |
|---|---|
| **Action** | Action |
| **WholesaleCompany** | BulkWaterSupplier |
| **ConsumptionAppliance** | Appliance |
| **ConsumptionProcess** | WaterUse |
| **ConsumptionProcess** | Activity |
| **WaterBody** | BodyOfWater |
| **WaterUtility** | WaterUtility |

| | |
|---|---|
| **GoverningBody** | Regulator |
| **occursAtEntity** | isModelOf |
| **AgriculturalConsumption** | AgriculturalUse |
| **SampledFeature** | SampledFeature |
| **SupplyChainProcess** | WaterResourceManagement |
| **TimeValuePair** | TimeValuePair |
| **Organisation** | WaterIndustry |
| **hasState** | hasState |
| **FeatureOfInterest** | FeatureOfInterest |
| **WaterAlert** | Alert |
| **StorageNode** | Storage |
| **TransformationNode** | Transformation |
| **DesignedArtifact** | Infrastructure |
| **WaterConsumer** | Consumer |
| **PhysicalArc** | Transport |
| **Property** | Phenomenon |
| **requiredToSolve** | solves |
| **WaterBalanceState** | State |
| **AbstractionNode** | Source |
| **observationResultTime** | hasTimestamp |
| **ObservationResult** | ObservationResult |
| **IndustrialConsumption** | IndustrialUse |
| **Observation** | Observation |
| **ConsumptionNode** | Sink |
| **SocialAgent** | Actor |
| **implementsConsumption** | peformsActivity |
| **PhysicalTopologicalEntity** | WaterResource |
| **isDescribedBy** | hasObservation |
| **hasValue** | hasValue |

**Table 51: Main heterogeneity between WatERP ontology and WISDOM**

| WatERP concept | Closest WISDOM concept(s) | Comments on Heterogeneity |
|---|---|---|
| TimeSeriesObservation | TimeSeriesStore, TimeValuePair | The WISDOM ontology perceives an observation result as the outcome of a single act of observation. This is such that an observation is made available at a single point in time, in accordance with SSN:ObservationResultTime. |
| Cluster | WaterConsumer subclasses | The WISDOM ontology models types of water consumers as subtypes of water consumer, as it is logically inconsistent to state that a consumer is a type of group of consumer, as a group must contain more than one member. |
| WaterUse | ConsumptionProcess | Whilst these classes appear weakly aligned, it is inconsistent to state that a group of consumers is a type of consumption process, which would be explicit in the WISDOM ontology if this alignment was formalised, unless the definition of the WaterUse class has been misunderstood. |
| Consumer | WaterConsumer, SocialAgent,Resident | The WatERP ontology states that a Consumer is a type of EndUser, but it is not clear which individuals may exist in the EndUser set but not in the Consumer set. In the WISDOM ontology, it is assumed that no such individuals exist, such that the two classes are |

| | | |
|---|---|---|
| | | equivalent. |
| UserClass | WaterConsumer, SocialAgent,Resident | The WatERP ontology states that a UserClass is a type of Consumer, but the WISDOM ontology models types of Consumer as subclasses. In order to model specific people or water consuming organisations, an individual of type WatERP:UserClass would be a class and an individual, which would require the use of OWL Full, as “a class can not also be an individual” [311]. |

**Table 52: Likely aligned WISDOM and IFC concepts**

| **WISDOM term** | **IFC term** |
|---|---|
| Sensor | IfcSensor |
| AlarmDevice | IfcAlarm |
| PhysicalQuantity | IfcPhysicalSimpleQuantity |
| Arc | IfcEdge |
| DesignedArtifact | IfcAsset |
| Pump | IfcPump |
| Building | IfcBuilding |
| Entity | IfcObject |
| PhysicalArtifact | IfcProduct |
| Process | IfcProcess |
| Node | IfcVertex |
| Fitting | IfcPipeFitting |
| Condition | IfcConstraint |
| Amount | IfcQuantityCount |
| Length | IfcQuantityLength |
| TopologicalNetworkEntity | IfcTopologicalRepresentationItem |

| | |
|---|---|
| PhysicalTopologicalEntity | IfcTopologicalRepresentationItem |
| Property | IfcProperty |
| TimeInterval | IfcQuantityTime |
| Area | IfcQuantityArea |
| Chamber | IfcDistributionChamberElement |
| WaterMeter | IfcFlowMeter |
| StorageAsset | IfcFlowStorageDevice |
| Weight | IfcQuantityWeight |
| hasLocation | IfcPlacement |
| TreatmentWorks | IfcFlowTreatmentDevice |
| Volume | IfcQuantityVolume |
| Main | IfcPipeSegment |
| Actuator | IfcActuator |
| Event | IfcEvent |
| TerminalNode | IfcFlowTerminal |
| Valve | IfcValve |
| ElectricAppliance | IfcElectricAppliance |
| WaterSensor | IfcFlowInstrument |
| Material | IfcMaterial |

**Table 53: Aligned terms between the WISDOM ontology and the INSPIRE utility data models**

| **WISDOM term** | **INSPIRE term** |
|---|---|
| WaterPipe | WaterPipe |
| Main | Pipe |
| WastePipe | SewerPipe |
| Node | Node |
| hasNetworkEntity | elements |
| hasInputArc | spokeEnd |
| isInNetwork | inNetwork |
| hasEndNode | endNode |
| PhysicalArc | UtilityLink |
| hasOutputArc | spokeStart |

| | |
|---|---|
| PhysicalNetwork | UtilityNetwork |
| Arc | Link |
| hasStartNode | startNode |
| PhysicalTopologicalEntity | Network::NetworkElement |
| PhysicalNode | Appurtenance |

## 8.2 APPENDIX B: SMART WATER ONTOLOGY INSTANCE MODELS

The input and output data for each instance created of the smart water ontology at the WISDOM pilot sites are now summarised.

7 CSV files were used as a basis for extracting information regarding the Cardiff pilot site; these described the system valves, meters, mains, control valves, boundary valves, hydrants, and asset sensors. In total, these represented 352 KB of data. The number of entities and properties in each of these sheets is summarised in Table 53 below.

**Table 54: Summary of Cardiff pilot input data**

| Entity type | Number of entities | Number of properties |
|---|---|---|
| **Asset sensor** | 29 | 5 |
| **Hydrants** | 269 | 24 |
| **Boundary valves** | 47 | 19 |
| **Control valves** | 12 | 16 |
| **Mains** | 1210 | 28 |
| **Meters** | 29 | 30 |
| **System valves** | 370 | 22 |

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 2183 KB in size, representing 17355 triples. This included the 1966 named entities, and the properties shown below in Table 54.

**Table 55: Summary of Cardiff pilot output knowledge base before inference**

| Entity type | Object | Datatype Properties |
|---|---|---|

| | Properties | |
|---|---|---|
| **Asset sensor** | atAsset, observes | comment, hasUnit |
| **Hydrants** | | hasIpid, hasXcoord, hasYcoord |
| **Boundary valves** | | hasNominalDiameter, hasUnit, hasIpid, hasXcoord, hasYcoord |
| **Control valves** | | hasNominalDiameter, hasUnit, hasIpid, hasXcoord, hasYcoord |
| **Mains** | hasMaterial, hasWaterType, hasSubtype | hasNominalDiameter, hasUnit, hasLength, isPumped, goesFromIpid, goesToIpid, hasIpid, hasAbsoluteDiameter |
| **Meters** | hasSubtype, observes | hasNominalDiameter, hasUnit, hasXcoord, hasYcoord, hasMeterType, hasMeterFunction, hasAttachedPipeType, hasIpid, hasSiteRef |
| **System valves** | | hasNominalDiameter, hasUnit, hasIpid, hasXcoord, hasYcoord |

6 CSV files were used as a basis for extracting information regarding the Tywyn and Aberdovey pilot site; these described the system valves, meters, mains, control valves, hydrants, and asset sensors. In total, these represented 572 KB of data. The number of entities and properties in each of these sheets is summarised in Table 55 below.

**Table 56: Summary of Tywyn and Aberdovey pilot input data**

| Entity type | Number of entities | Number of properties |
|---|---|---|
| **Asset sensor** | 29 | 5 |
| **Hydrants** | 261 | 18 |
| **Control valves** | 46 | 18 |
| **Mains** | 1517 | 19 |
| **Meters** | 26 | 15 |
| **System valves** | 485 | 19 |

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 2142 KB in size, representing 21122 triples. This included the 2363 named entities, and the properties shown below in Table 56.

**Table 57: Summary of Tywyn and Aberdovey pilot output knowledge base before inference**

| Entity type | Object Properties | Datatype Properties |
|---|---|---|
| **Asset sensor** | atAsset, observes | comment, hasUnit |
| **Hydrants** | | hasIpid, hasXcoord, hasYcoord |
| **Control valves** | | hasNominalDiameter, hasUnit, hasIpid, hasXcoord, hasYcoord |
| **Mains** | hasMaterial, hasWaterType, hasSubtype | hasNominalDiameter, hasUnit, hasLength, isPumped, goesFromIpid, goesToIpid, hasIpid, hasAbsoluteDiameter |
| **Meters** | hasSubtype, observes | hasNominalDiameter, hasUnit, hasXcoord, hasYcoord, hasMeterType, hasMeterFunction, hasAttachedPipeType, hasIpid, hasSiteRef |
| **System valves** | | hasNominalDiameter, hasUnit, hasIpid, hasXcoord, hasYcoord |

As the Gowerton pilot site only included waste water assets, its knowledge base is significantly different to the Cardiff and Tywyn & Aberdovy sites. 5 CSV files were used as a basis for extracting information regarding the Gowerton pilot site; these described the conduits, nodes, pumps, sensors and subcatchments. In total, these represented 5.28 MB of data. The number of entities and properties in each of these sheets is summarised in Table 57 below.

**Table 58: Summary of Gowerton pilot input data**

| Entity type | Number of entities | Number of properties |
|---|---|---|
| **Sensors** | 103 | 12 |
| **Subcatchments** | 1372 | 140 |

| | | |
|---|---|---|
| **Conduits** | 2574 | 204 |
| **Nodes** | 2621 | 71 |
| **Pumps** | 40 | 186 |

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 6582 KB in size, representing 60081 triples. This included the 6674 named entities, and the properties shown below in Table 58.

**Table 59: Summary of Gowerton pilot output knowledge base before inference**

| Entity type | Object Properties | Datatype Properties |
|---|---|---|
| **Sensors** | atAsset, observes, subtype | hasXcoord, hasYcoord |
| **Subcatchments** | subtype | has Area, hasXcoord, hasYcoord, hasPopulation |
| **Conduits** | Subtype, hasMaterial, hasUpstreamNode, hasDownstreamNode | hasNominalDiameter, hasXcoord, hasYcoord, hasLength, hasPipeShape, hasWidth, hasHeight, hasBottomRoughness, hasTopRoughness, hasUpstreamInvertlevel, hasUpstreamHeadloss, hasDownstreamInvertlevel, hasDownstreamHeadloss, hasGradient, hasCapacity |
| **Nodes** | Subtype, atAsset | hasXcoord, hasYcoord |
| **Pumps** | Subtype, WaterType, | PumpComment, hasOnDelay, hasOffDelay, hasDischarge |

The Italian Pilot site was instantiated using a Python script to extract knowledge from the hydraulic model developed of the Italian pilot site, using the RDFlib Python library previously described and the EPANETTOOLS Python library. This meant that instead of the Welsh pilot method of exporting CSV files, then parsing the CSV files into Python objects, the EPANET model could be parsed directly into Python objects, and then iterated over to add statements to an RDFlib graph and the WISDOM namespace, before being serialized into an RDF file. The Italian pilot input data was split across several sections of an EPANET input file, which described the 426 entities within 42 KB of data, as detailed in Table 59.

**Table 60: Summary of Italian pilot input data**

| Entity | Number of entities | Properties |
|---|---|---|
| **Junctions** | 99 | Elevation, demand, X-Coord, Y-Coord |
| **Pipes** | 99 | Node1, Node2, Length, Diameter, Roughness, Status |
| **Pipe vertices** | 190 | Pipe, X-Coord, Y-Coord |
| **Pumps** | 19 | Node1, Node2, Parameters |
| **Reservoirs** | 19 | Head |

Following the production of the knowledge base, an RDF/XML file was produced, which included only the Abox triples, and was 222 KB in size, representing 1942 triples. This included the 426 named entities, and the properties shown below in Table 60.

**Table 61: Summary of Italian pilot output knowledge base before inference**

| Entity | Number of entities | Properties |
|---|---|---|
| **Junctions** | 99 | Elevation, demand, X-Coord, Y-Coord |
| **Pipes** | 99 | hasStartNode, hasEndNode , Length, Diameter, Roughness |
| **Pipe vertices** | 190 | Pipe, X-Coord, Y-Coord |
| **Pumps** | 19 | hasStartNode, hasEndNode |
| **Reservoirs** | 19 | |

## 8.3 APPENDIX C: WATER ONTOLOGY SWRL RULES

A full description of the SWRL rules embedded in the smart water ontology is provided here. For each rule, the inferred property is declared as a heading, followed by a human readable description, and finally the formal SWRL rule.

1. deployedAt

As sensors are not explicitly described in terms of the node which they are deployed at, this is fundamental knowledge which must be inferred in order to contextualise the capability of the deployed sensors.

SENSOR(?S) ^ ATASSET(?S, ?A) ^ TOPOLOGICALNETWORKENTITY(?E) ^ ATASSET(?E, ?A) -> DEPLOYEDATENTITY(?S, ?E)

2. observes

Whilst it may be explicitly stated in the knowledge base that a sensor observes a certain property, this may not be directly stated, as the SSN modelling pattern is that a sensor has a measurement capability, which is then for a certain property. If this is the case, it is useful to infer the direct link between the sensor and the property, to facilitate the previous rules.

HASMEASUREMENTCAPABILITY(?S, ?MC) ^ FORPROPERTY(?MC, ?P) -> OBSERVES(?S, ?P)

3. isActive

If an alert has an acceptable range, and it is triggered by a sensor, and the sensor’s latest reading falls outside of the acceptable range, the alarm is triggered. This can occur if the reading is greater than the allowable maximum, or smaller than the allowable minimum, otherwise the alarm is not active.

WATERALERT(?A) ^ HASALERTCONDITION(?A, ?AC1) ^ HASACCEPTABLERANGE(?AC1, ?AR) ^ HASMAXVALUE(?AR, ?XMAX) ^ SENSOR(?S) ^ TRIGGERSALERT(?S, ?A) ^ HASLATESTOUTPUT(?S, ?TVP) ^ HASVALUE(?TVP, ?X) ^ SWRLB:GREATERTHAN(?X, ?XMAX) -> ISACTIVE(?A, TRUE)

WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^ HASMINVALUE(?AR,?XMIN)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:LESSTHAN(?X,?XMIN)-> ISACTIVE(?A,TRUE)

WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^ HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT

(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:LESSTHANOREQUAL(?X,?XMAX)^SWRLB:GREATERTHANOREQUAL(?X,?XMIN)-> ISACTIVE(?A,FALSE)

4. hasDownstreamEntity

In order to generalise pipes, pumps, and reservoirs etc. to determine what is upstream or downstream of an entity, it is useful to use the IPID values held in the legacy GIS database to infer knowledge about flow chronology through the entities. This allows later inference of whether an entity is affected by any given problem, and greatly simplifies those rules.

If an entity goes from an entity with IPID of i, and another entity has an IPID of i, then the latter is downstream of the former, and vice versa.

GOESFROMIPID(?P, ?I) ^ HASIPID(?U, ?I) -> HASUPSTREAMENTITY(?P, ?U) ^ HASDOWNSTREAMENTITY(?U, ?P)

GOESTOIPID(?P,?I)^HASIPID(?D, ?I) -> HASUPSTREAMENTITY(?D, ?P) ^ HASDOWNSTREAMENTITY(?P, ?D)

5. hasProblem

In the situation that an alert is active, and the alert is caused by a certain problem, it is beneficial to connect the problem entity to the problem directly, which this rule achieves.

If an entity has an active alert, and the alert is for a certain problem, the entity has that problem.

HASALERT(?E, ?A) ^ FORPROBLEM(?A, ?P) ^ISACTIVE(?A,TRUE) -> HASPROBLEM(?E, ?P)

6. hasAffectedEntity

Tracing the impact of a problem downstream in a water network to determine further problems which the problem could cause, and its negative consequences for customers, is both highly beneficial and challenging, due to the system complexity. Further, if a problem is reported at a downstream entity, one expert task is tracing backwards in the value chain to determine if an upstream problem could be causing it. This rule aims to empower water experts by telling them upfront if an entity is affected by any upstream problems. Further, the knowledge of an entity being affected by an upstream problem could be used automatically by a further rule to infer a likely problem at that entity, and even a required action to proactively mitigate the overall impact of the initial problem. Note that ‘hasProblemEntity’ is a sub-property of ‘hasAffectedEntity’, allowing this inference to propagate to all downstream elements.

If a water alert is active, and the problem entity has a downstream entity, then that entity is affected by the alert.

WATERALERT(?A) ^ ISACTIVE(?A,TRUE) ^ HASAFFECTEDENTITY(?A, ?E) ^ HASDOWNSTREAMENTITY(?E, ?D) -> HASAFFECTEDENTITY(?A, ?D)

### *7.* AFFECTEDBYPROBLEM

This rule continues the benefits of rule 5 by directly linking the downstream entity with the problem which it is affected by. Note that ‘hasProblem’ is a sub-property of ‘affectedByProblem’, allowing the inference to propagate to all downstream entities.

If an entity has a downstream entity and is affected by a problem, then that downstream entity is also affected by the problem.

HASDOWNSTREAMENTITY(?E, ?D) ^ AFFECTEDBYPROBLEM(?E, ?P) -> AFFECTEDBYPROBLEM(?D, ?P)

### 8. hasSeverity

This inference ability explores the possibility of evaluating how severe a problem is, based on how far outside the acceptable range a current sensor reading is. This has been achieved is a simple manner by finding the relative distance which the reading is outside of the acceptable range. This could be explored further with more specific use cases, and more domain knowledge about the criteria for problem severity, such as the likely total future impact on the organization’s KPIs. However, the current approach shows that the knowledge-based approach allows further knowledge to be derived quite easily about the current situation, to empower decision makers without them having to do their own analysis of the data.

Firstly, in the case of the reading being greater than the maximum allowable value, severity is defined by equation 1.

$$\text{IF VAL}_{\text{ACTUAL}} > \text{VAL}_{\text{MAX}}:\ \text{SEVERITY} = \frac{\text{VAL}_{\text{ACTUAL}} - \text{VAL}_{\text{MAX}}}{\left(\frac{\text{VAL}_{\text{MAX}} - \text{VAL}_{\text{MIN}}}{2}\right)} \tag{1}$$

If a problem causes an alert and the alert has an acceptable range, and the alert is triggered by a sensor whose latest reading is above that range, the severity of the problem is calculated as per equation 1.

PROBLEM(?P)^ISCAUSEOFALERT(?P,?A)^WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:GREATERTHAN(?X,?XMAX)^SWRLB:SUBTRACT(?X1,?XMAX,?XMIN)^SWRLB:DIVIDE(?X2,?X1,2)^SWRLB:SUBTRACT(?SEVABS,?X,?XMAX)^SWRLB:DIVIDE(?SEVREL,?SEVABS,?X2)->HASSEVERITY(?P,?SEVREL)

In parallel to the above rule, the opposite logic holds if the sensor reading is below the minimum acceptable range, where the severity is calculated as per equation 2 below.

$$\text{IF VAL}_{\text{ACTUAL}} < \text{VAL}_{\text{MIN}}:\ \text{SEVERITY} = \frac{\text{VAL}_{\text{MIN}} - \text{VAL}_{\text{ACTUAL}}}{\left(\frac{\text{VAL}_{\text{MAX}} - \text{VAL}_{\text{MIN}}}{2}\right)} \tag{2}$$

PROBLEM(?P)^ISCAUSEOFALERT(?P,?A)^WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:LESSTHAN(?X,?XMIN)^SWRLB:SUBTRACT(?X1,?XMAX,?XMIN)^SWRLB:DIVIDE(?X2,?X1,2)^SWRLB:SUBTRACT(?SEVABS,?XMIN,?X)^SWRLB:DIVIDE(?SEVREL,?SEVABS,?X2)->HASSEVERITY(?P,?SEVREL)

If the sensor’s reading is within the acceptable range, then the severity of the problem is 0.

PROBLEM(?P)^ISCAUSEOFALERT(?P,?A)^WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVAL

UE(?TVP,?X)^SWRLB:LESSTHANOREQUAL(?X,?XMAX)^SWRLB:GREATERTHANOREQUAL(?X,?XMIN)->HASSEVERITY(?P,0.0)

9. hasDetectionTime

Given that the knowledge base will be iteratively updated as new sensor readings are received, and alerts may not be observed immediately, it would be beneficial to inform decision makers exactly when a problem was first observed. This is achieved by noting the time at which the sensors latest reading is outside the acceptable range, but the sensor's previous reading was inside the acceptable range.

If an alert has an acceptable range, and is triggered by a sensor, and the sensor's latest reading is outside that range, but its previous reading was inside the range, then the detection time of the problem is the latest reading's timestamp. This can occur when the reading is above the maximum range, or below the minimum range.

PROBLEM(?P)^ISCAUSEOFALERT(?P,?A)^WATERALERT(?A)^HASALERTCONDITION(?A,?AC1)^HASACCEPTABLERANGE(?AC1,?AR)^HASMINVALUE(?AR,?XMIN)^HASMAXVALUE(?AR,?XMAX)^SENSOR(?S)^TRIGGERSALERT(?S,?A)^HASLATESTOUTPUT(?S,?TVP)^HASVALUE(?TVP,?X)^SWRLB:GREATERTHAN(?X,?XMAX)^HASTIMESTAMP(?TVP,?TIME)^HASPREVIOUSOUTPUT(?S,?TVPPREV)^DIFFERENTFROM(?TVP,?TVPPREV)^HASVALUE(?TVPPREV,?XPREV)^SWRLB:LESSTHANOREQUAL(?XPREV,?XMAX)^SWRLB:GREATERTHANOREQUAL(?XPREV,?XMIN)->HASDETECTIONTIME(?P,?TIME)

PROBLEM(?P) ^ ISCAUSEOFALERT(?P, ?A) ^ WATERALERT(?A) ^ HASALERTCONDITION(?A, ?AC1) ^ HASACCEPTABLERANGE(?AC1, ?AR) ^ HASMINVALUE(?AR, ?XMIN) ^ HASMAXVALUE(?AR, ?XMAX) ^ SENSOR(?S) ^ TRIGGERSALERT(?S, ?A) ^ HASLATESTOUTPUT(?S, ?TVP) ^ HASVALUE(?TVP, ?X) ^ SWRLB:LESSTHAN(?X, ?XMIN) ^ HASTIMESTAMP(?TVP, ?TIME) ^ HASPREVIOUSOUTPUT(?S, ?TVPPREV) ^ DIFFERENTFROM(?TVP, ?TVPPREV) ^ HASVALUE(?TVPPREV, ?XPREV) ^ SWRLB:LESSTHANOREQUAL(?XPREV, ?XMAX) ^ SWRLB:GREATERTHANOREQUAL(?XPREV, ?XMIN) -> HASDETECTIONTIME(?P, ?TIME)

## 8.4 APPENDIX D: FULL SMART CITY SERVER API SPECIFICATION

The full API for the smart city server developed in the 3rd stage of the investigation is detailed through Table 61 - Table 77

**Table 62: GET method API for SPARQL endpoint of smart city server**

<table>
<tr><th colspan="3">GET /sparql/{dataset}/query</th></tr>
<tr><th colspan="3">Description</th></tr>
<tr><td colspan="3">Endpoint for retrieving information from the knowledge base through SELECT, CONSTRUCT, ASK, and DESCRIBE SPARQL queries, passed as URL encoded strings through the query parameter.</td></tr>
<tr><th colspan="3">Request parameters</th></tr>
<tr><td>query</td><td>SPARQL query string</td><td>The GET method can only be used to retrieve information. Other methods must be used to update, insert, or delete data.</td></tr>
<tr><th colspan="3">Response</th></tr>
<tr><td>String, CSV, RDF/XML, Boolean or JSON</td><td colspan="2">Response type varies depending on the SPARQL query form. SELECT returns a string by default, CONSTRUCT and DESCRIBE return an RDF/XML graph, and ASK returns a Boolean.</td></tr>
</table>

**Table 63: POST method API for the KairosDB endpoint of the smart city server**

<table>
<tr><th colspan="2">POST /data</th></tr>
<tr><th colspan="2">Description</th></tr>
<tr><td colspan="2">A KairosDB query is passed to the endpoint in the body of the request as a JSON string specifying the ID of the sensor, as well as the desired date range, aggregation, time zone, grouping, return order, and maximum number of returned points. The query can also be executed with the GET method by encoding the query and passing it to the ‘query’ parameter.</td></tr>
<tr><th colspan="2">Response</th></tr>
<tr><td>JSON</td><td>The response is a JSON object containing a hierarchy of JSON objects detailing the number of data points returned, the sensor ID returned, other echoes of the query processed, and finally the data points.</td></tr>
</table>

**Table 64: API specification for the Hypercat root endpoint**

| GET /cat |
|---|
| Description |

| | | |
|---|---|---|
| Top level endpoint for retrieving the overall catalogue of resources available through the server. This endpoint masks the complexity of SPARQL through a binding between the Smart City Ontology developed and the Hypercat specification. Currently, only devices offering web services are exposed through the Hypercat API, although this could easily be adjusted to suit client requirements. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | The response is a JSON object complying with the BSI:PAS 212 specification of Hypercat catalogues. | |

**Table 65: API specification for the Hypercat item endpoint of the smart city server**

| GET /cat/{item_ID} | | |
|---|---|---|
| Description | | |
| Item level endpoint for retrieving information about a specific Hypercat item. Currently, only devices offering web services are represented as Hypercat Items, but this could be adjusted to suit client requirements, such as also including buildings, or any set of named individuals from the triple store. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | The response is a JSON object complying with the BSI:PAS 212 specification of Hypercat item descriptions. | |

**Table 66: API specification for the Hypercat item description endpoint of the smart city server**

| GET /cat/{item_ID}/description | | |
|---|---|---|
| Description | | |
| Retrieves the name and human-readable description of a Hypercat item. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | JSON object with two parameters: name and description, the values of which are Strings. | |

**Table 67: Top level endpoint of the BIM interface of the smart city server**

| GET /bim | | |
|---|---|---|
| Description | | |
| Top level endpoint of the bim interface, which provides a name, description, and size of all the BIM models stored in the server. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | JSON object containing an array of objects, where each object contains three parameters: name, description, and size, the first two of which have String values, and the latter of which has an integer value for the number of Kilobytes of data. | |

**Table 68: Building level endpoint of the BIM interface of the smart city server**

| GET /bim/{id} | | |
|---|---|---|
| Description | | |
| Top level endpoint for a building. Returns the name and description of the building described in the relevant IFC file. | | |
| Request parameters | | |
| id | String | The ID of the building which is the subject of the query |
| Response | | |
| JSON | JSON object three parameters: name, description, and size, the first two of which have String values, and the latter of which has an integer value for the number of Kilobytes of data. | |

**Table 69: Building level source file endpoint of the BIM interface of the smart city server**

| GET /bim/{id}/src | | |
|---|---|---|
| Description | | |
| Returns the STEP source code of the building from the relevant IFC file. | | |
| Request parameters | | |
| id | String | The ID of the building which is the subject of the query |
| Response | | |

| | |
|---|---|
| String | Returns the contents of the IFC file as a string |

**Table 70: Entity level endpoint of the BIM interface of the smart city server**

| GET /bim/{id}/{guid} | | |
|---|---|---|
| Description | | |
| Entity level endpoint of the BIM interface, which returns details about the element specified by the GUID. | | |
| Request parameters | | |
| id | String | The ID of the building which is the subject of the query |
| guid | String | The ID of the element which is the subject of the query |
| Response | | |
| JSON | JSON object containing an array of objects, where each object represents a property, and has a number of parameters depending on the property. | |

**Table 71: Property level endpoint of the BIM interface of the smart city server**

| GET /bim/{id}/{guid}/{property} | | |
|---|---|---|
| Description | | |
| Entity property interface of the BIM interface, which returns the property of the element specified by the GUID and property name. | | |
| Request parameters | | |
| id | String | The ID of the building which is the subject of the query |
| guid | String | The ID of the element which is the subject of the query |
| property | String | The name of the property to be retrieved for the subject element |
| Response | | |
| JSON | JSON object containing the name of the property and the value of the property, where the value type depends on the property | |

**Table 72: Top level endpoint of the CityGML interface of the smart city server**

| GET /citygml |
|---|
| Description |

| | | |
|---|---|---|
| Top level endpoint; returns the names, sizes and descriptions of all the CityGML files stored on the server, where available. | | |
| Request parameters | | |
| None | | |
| Response | | |
| JSON | JSON object containing an array of objects, where each object contains three parameters: name, description, and size, the first two of which have String values, and the latter of which has an integer value for the number of Kilobytes of data. | |

**Table 73: Model level endpoint of the CityGML interface of the smart city server**

| GET /citygml/{id} | | |
|---|---|---|
| Description | | |
| City model level endpoint, which returns the name, description, and size of the city model specified, if available. | | |
| Request parameters | | |
| id | String | ID string of the CityGML model |
| Response | | |
| JSON | JSON object containing three parameters: name, description, and size, the first two of which have String values, and the latter of which has an integer value for the number of Kilobytes of data. | |

**Table 74: Model level source endpoint of the CityGML interface of the smart city server**

| GET /citygml/{id}/src | | |
|---|---|---|
| Description | | |
| Machine-readable city model level endpoint, which returns the source code of the city model specified, if available. | | |
| Request parameters | | |
| id | String | ID string of the CityGML model |
| Response | | |
| XML | Returns the CityGML source code for the model specified | |

**Table 75: Entity level endpoint of the CityGML interface of the smart city server**

| GET /citygml/{id}/{gmlid} | | |
|---|---|---|
| Description | | |
| Entity-level API which returns the properties of a CityGML entity, specified by the ID of the model and entity. | | |
| Request parameters | | |
| id | String | ID string of the CityGML model |
| gmlid | String | ID string of the CityGML entity |
| Response | | |
| JSON | JSON object containing an array of objects, where each object represents a city object, and has a number of parameters depending on the object. | |

**Table 76: Entity level source endpoint of the CityGML interface of the smart city server**

| GET /citygml/{id}/{gmlid}/src | | |
|---|---|---|
| Description | | |
| Entity level source code API, which returns the XML description of a CityGML entity, specified by the ID of the model and entity. | | |
| Request parameters | | |
| id | String | ID string of the CityGML model |
| gmlid | String | ID string of the CityGML entity |
| Response | | |
| XML | Returns the CityGML source code for the entity specified | |

**Table 77: Root endpoint of the smart city server, for the GUI**

| GET / | | |
|---|---|---|
| Description | | |
| 3D graphical interface for human exploration of the available resources and systems in the pilot site, based on cesium.js. | | |
| Request parameters | | |
| None | | |
| Response | | |
| HTML | HTML page containing the Cesium.js widget zoomed to the pilot site data | |

**Table 78: Object level endpoint of the GUI for resolving object URIs to human-readable information**

| GET /{id} | | |
|---|---|---|
| Description | | |
| 3D graphical interface for human exploration of the available resources and systems in the pilot site, based on cesium.js. By adding the ID of an object after the root of the server, it initialises focused on that object, displaying its information box. | | |
| Request parameters | | |
| id | String | ID of the object to be focused on |
| Response | | |
| HTML | HTML page containing the Cesium.js widget zoomed to the pilot site data | |

## 8.5 APPENDIX E: RESEARCH MOTIVATION

After completing 3 years of higher education towards a Master's degree in Civil Engineering at Cardiff University, I had gained a grounding in many theories and skills relevant to the lifecycle of built environments. However, I observed that the value of emerging technologies was poorly appreciated in the AECFM industry, and many others which exhibit similar inertia and are very risk-adverse. I realised that continuing to pursue traditional approaches would not sufficiently address the growing pressures faced by the complex system of systems present in urban environments, from economic as well as environmental perspectives. The potential impact of certain technologies, and the prospect of contributing to their development and implementation greatly motivated me, and so I decided to pursue a BEng + PhD route instead, as this would better prepare me with skills in leading edge technologies and independent investigation.

Regarding my choice of research topic, the Internet of Things, artificial intelligence, and the semantic web, have all been foreshadowed as having the potential to revolutionise the modern world, yet the intersection of these fields has barely been considered. The value proposition of their convergence is unprecedented, and wielded in the right manner could even herald a bright new age of humanity. Yet realising these exceptional claims carries exceptional challenges. My motivation for investigating this space is to contribute to the emerging global effort of overcoming these obstacles. My intention is to produce and disseminate knowledge and artefacts through collaborative innovation, alongside experts in various fields, through which I also aim to gain a breadth of expertise across these fields, with a depth of expertise at their intersection. By adopting a pragmatic stance and focusing on the value proposition of this complex research space, I hope to move this global effort a small step closer to achieving the vision of a truly intelligent web of things.

## 8.6 APPENDIX F: INTENDED AUDIENCE

Whilst this thesis is targeted at an academic audience with at least some background in the central topics (semantic technologies, AI, and IoT), an effort has been made to achieve a broader appeal by considering the research and results from various viewpoints. Especially, the relevance of the work to businesses and governmental decision makers was strongly considered. The nature of a PhD thesis is necessarily more verbose and focused on providing a *knowledge* contribution than may appeal to such audiences, so a shorter alternative version will be made available. Interested persons, and anyone with questions, are encouraged to contact the author by email at shaunkhowell@gmail.com.